GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2001, 15:09:06 ; Search time 1481.25 Seconds
(without alignments)
8990.874 Million cell updates/sec

Title: US-09-308-080-1
Perfect score: 861
Sequence: 1 TGTTAATGAAGATAAATATT.....AGTGGGAAATAATTTATTAA 861

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	773.6	89.8	180286	80	AL356574	AL356574 Homo sapi
	2	607	70.5	626	97	HSU57655	U57655 Human dihyd
	3	486.6	56.5	854	93	HSDYPDGEN	X95670 H.sapiens D
	4	172.8	20.1	3951	97	HSU09178	U09178 Human dihyd
	5	172.8	20.1	3957	9	AR027740	AR027740 Sequence
	6	171.2	19.9	3078	85	AB003063	AB003063 Homo sapi
	7	171.2	19.9	4409	97	HSU20938	U20938 Human lymph
	8	154.8	18.0	4421	7	BTU20981	U20981 Bos taurus

	9	151.6	17.6	4441	7	SSU09179 U09179 Sus scrofa
	10	151.6	17.6	4447	9	AR027741 AR027741 Sequence
	11	146.8	17.0	4358	94	D85035 D85035 Rattus norv
	12	72	8.4	43803	6	CELC25F6 U39742 Caenorhabdi
	13	71.4	8.3	24429	63	AC014157 AC014157 Drosophil
c	14	71.4	8.3	303092	4	AE003446 AE003446 Drosophil
	15	69.8	8.1	2348	6	DMU65491 U65491 Drosophila
	16	56.8	6.6	117674	84	DMBR30C13 AL122025 Drosophil
c	17	51	5.9	141953	71	AC044793 AC044793 Homo sapi
c	18	50	5.8	156060	60	AC004153 AC004153 Plasmodiu
	19	49.8	5.8	150803	61	AC010228 AC010228 Homo sapi
	20	49.6	5.8	1017	6	DDCPRO2U X04775 Dictyosteli
c	21	49.4	5.7	147490	90	AL445207 AL445207 Human DNA
	22	49.4	5.7	160728	70	AC026816 AC026816 Homo sapi
c	23	49.4	5.7	161025	66	AC021276 AC021276 Homo sapi
	24	49.4	5.7	164450	82	AP000794 AP000794 Homo sapi
	25	49.2	5.7	168674	69	AC025645 AC025645 Homo sapi
c	26	48.8	5.7	94536	15	T14P8 AF069298 Arabidops
	27	48.8	5.7	194892	13	ATCHRIV6 AL161494 Arabidops
c	28	48.8	5.7	217327	79	AL162723 AL162723 Homo sapi
c	29	48.2	5.6	1149	5	AF044859 AF044859 Eurema da
c	30	48.2	5.6	12029	4	AE001412 AE001412 Plasmodiu
	31	48	5.6	590	96	PARMTDI4D K00907 paramecium
	32	48	5.6	692	96	PARMTDI4H K00911 paramecium
	33	48	5.6	726	96	PARMTDI4C K00906 paramecium
	34	48	5.6	760	96	PARMTDI4A K00904 paramecium
	35	48	5.6	40469	6	MIPAGEN X15917 Paramecium
	36	48	5.6	47573	5	AF030694 AF030694 Plasmodiu
c	37	47.4	5.5	112519	85	AB041992 AB041992 Homo sapi
	38	47.4	5.5	148965	72	AC060784 AC060784 Homo sapi
	39	47.4	5.5	165823	71	AC027625 AC027625 Homo sapi
c	40	47.4	5.5	177920	64	AC016700 AC016700 Homo sapi
c	41	47.4	5.5	300000	91	AP002529 AP002529 Homo sapi
e	42	47.2	5.5	86827	96	PFMAL3P5 AL034556 Plasmodiu
c	43	47	5.5	7218	10	I66494 I66494 Sequence 14
	44	47	5.5	65691	96	PFMAL3P1 Z97348 Plasmodium
	45	46.8	5.4	87509	12	AC004482 AC004482 Arabidops

ALIGNMENTS

```
RESULT   1
AL356574/c
LOCUS       AL356574    180286 bp    DNA              HTG       23-JAN-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-359C24, *** SEQUENCING IN
            PROGRESS ***, 26 unordered pieces.
ACCESSION   AL356574
VERSION     AL356574.3  GI:9797568
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180286)
  AUTHORS   Plumb,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 12, 2000 this sequence version replaced gi:9213933.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA359C24
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 166105 bases at least Q40
            Consensus quality: 171672 bases at least Q30
            Consensus quality: 174563 bases at least Q20
            Insert size: 177786; sum-of-contigs
            Insert size: 186027; 5.1% error; agarose-fp
            Quality coverage: 3.18x in Q20 bases; sum-of-contigs Quality
            coverage: 3.10x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    15090: contig of 15090 bp in length
            *    15091 15190: gap of      100 bp
            *    15191    23545: contig of 8355 bp in length
            *    23546 23645: gap of      100 bp
            *    23646    26766: contig of 3121 bp in length
            *    26767 26866: gap of      100 bp
            *    26867    38453: contig of 11587 bp in length
            *    38454 38553: gap of      100 bp
            *    38554    45878: contig of 7325 bp in length
            *    45879 45978: gap of      100 bp
            *    45979    51984: contig of 6006 bp in length
            *    51985 52084: gap of      100 bp
            *    52085    58204: contig of 6120 bp in length
            *    58205 58304: gap of      100 bp
            *    58305    71630: contig of 13326 bp in length
            *    71631 71730: gap of      100 bp
            *    71731    75338: contig of 3608 bp in length
            *    75339 75438: gap of      100 bp
            *    75439    85213: contig of 9775 bp in length
            *    85214 85313: gap of      100 bp
            *    85314    88517: contig of 3204 bp in length
            *    88518 88617: gap of      100 bp
            *    88618    92454: contig of 3837 bp in length
            *    92455 92554: gap of      100 bp
            *    92555    98200: contig of 5646 bp in length
            *    98201 98300: gap of      100 bp
            *    98301   111029: contig of 12729 bp in length
            *   111030 111129: gap of      100 bp
            *   111130   117311: contig of 6182 bp in length
            *   117312 117411: gap of      100 bp
            *   117412   121093: contig of 3682 bp in length
            *   121094 121193: gap of      100 bp
            *   121194   137336: contig of 16143 bp in length
            *   137337 137436: gap of      100 bp
            *   137437   141486: contig of 4050 bp in length
            *   141487 141586: gap of      100 bp
            *   141587   146703: contig of 5117 bp in length
            *   146704 146803: gap of      100 bp
            *   146804   149034: contig of 2231 bp in length
            *   149035 149134: gap of      100 bp
            *   149135   154764: contig of 5630 bp in length
            *   154765 154864: gap of      100 bp
            *   154865   159990: contig of 5126 bp in length
            *   159991 160090: gap of      100 bp
            *   160091   163742: contig of 3652 bp in length
            *   163743 163842: gap of      100 bp
            *   163843   166322: contig of 2480 bp in length
            *   166323 166422: gap of      100 bp
            *   166423   176700: contig of 10278 bp in length
            *   176701 176800: gap of      100 bp
            *   176801   180286: contig of 3486 bp in length.
FEATURES             Location/Qualifiers
     source          1. .180286
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-359C24"
                     /clone_lib="RPCI-11.2"
```

```
     misc_feature    1. .15090
                     /note="assembly_fragment:00687
                     clone_end:SP6
                     vector_side:left"
     misc_feature    15191. .23545
                     /note="assembly_fragment:00038
                     fragment_chain:1"
     misc_feature    23646. .26766
                     /note="assembly_fragment:00972
                     fragment_chain:1"
     misc_feature    26867. .38453
                     /note="assembly_fragment:01030
                     fragment_chain:1"
     misc_feature    38554. .45878
                     /note="assembly_fragment:00178
                     fragment_chain:2"
     misc_feature    45979. .51984
                     /note="assembly_fragment:00498
                     fragment_chain:2"
     misc_feature    52085. .58204
                     /note="assembly_fragment:00217
                     fragment_chain:3"
     misc_feature    58305. .71630
                     /note="assembly_fragment:01029
                     fragment_chain:3"
     misc_feature    71731. .75338
                     /note="assembly_fragment:00265
                     fragment_chain:4"
     misc_feature    75439. .85213
                     /note="assembly_fragment:01189
                     fragment_chain:4"
     misc_feature    85314. .88517
                     /note="assembly_fragment:00320
                     fragment_chain:5"
     misc_feature    88618. .92454
                     /note="assembly_fragment:00693
                     fragment_chain:5"
     misc_feature    92555. .98200
                     /note="assembly_fragment:00834
                     fragment_chain:6"
     misc_feature    98301. .111029
                     /note="assembly_fragment:00385
                     fragment_chain:6"
     misc_feature    111130. .117311
                     /note="assembly_fragment:00932
                     fragment_chain:7"
     misc_feature    117412. .121093
                     /note="assembly_fragment:01129
                     fragment_chain:7"
     misc_feature    121194. .137336
                     /note="assembly_fragment:00160"
     misc_feature    137437. .141486
                     /note="assembly_fragment:00194"
     misc_feature    141587. .146703
                     /note="assembly_fragment:00234"
     misc_feature    146804. .149034
                     /note="assembly_fragment:00529"
     misc_feature    149135. .154764
                     /note="assembly_fragment:00590"
     misc_feature    154865. .159990
                     /note="assembly_fragment:01001"
     misc_feature    160091. .163742
                     /note="assembly_fragment:01119"
     misc_feature    163843. .166322
                     /note="assembly_fragment:01294"
     misc_feature    166423. .176700
                     /note="assembly_fragment:01323"
     misc_feature    176801. .180286
                     /note="assembly_fragment:01326"
BASE COUNT    57303 a  32087 c  31562 g  56812 t   2522 others
ORIGIN
```

```
  Query Match             89.8%;  Score 773.6;  DB 80;  Length 180286;
  Best Local Similarity   95.7%;  Pred. No. 2.2e-141;
  Matches  827;  Conservative    0;  Mismatches   34;  Indels    3;  Gaps    3;

Qy       1 TGTTAATGAAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGT 60
           ||||||||| |||||||||||| ||||| |||| ||||||||||| ||||||||||
Db   50426 TGTTAATGAATATAAATATTTTTATTTTTGCGCTATTCTAAACCTAGAATTACAAGAAGT 50367

Qy      61 AATTTATCTGGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTT 120
           |||||| ||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db   50366 AATTTAGCTAAAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTT 50307

Qy     121 CAATTCTAATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATA 180
           |||||||||||||||||||||||||||||||||||||| ||||||| ||||||||||||
Db   50306 CAATTCTAATTTAATGTATATTAAAAATTCCTCTGCAAAAATGTGAGAAGGGACCTCATA 50247

Qy     181 AAATATTGTCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50246 AAATATTGTCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGG 50187

Qy     241 AGACTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCAT 300
           |||||||||||||||||||||||| |||||||||||||||||||||||| ||||||
Db   50186 AGACTCAATATCTTTACTCTTTCATCAGGACATTGTGACAAATGTTTCCCCCAGAATCAT 50127

Qy     301 CCGGGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50126 CCGGGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGA 50067

Qy     361 GCTCATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50066 GCTCATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGA 50007

Qy     421 CTTCCCAGACAACaTAAGTGTGA-TAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGT 479
           ||| ||||||||| |||||||| | || ||||||||||||||||||||||||||||||
Db   50006 CTTTCCAGACAACGTAAGTGTGATTTAACATCTAAAACAAGAGAATTGGCATAAGTTGGT 49947

Qy     480 GAATGTTTATTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCA 539
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   49946 GAATGTTTATTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATTA 49887

Qy     540 ACGAATCTGCCAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAG 599
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   49886 AAGAATCTGCCAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAG 49827

Qy     600 AACTCATAAAAACCCACACAATGTGAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGA 659
           |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   49826 AACTCATAAAAACCCACACAATGTGAAGCTTTGTTATAAATGGGTGCCATGTAAGATGGA 49767

Qy     660 AGAAGTATCTACATAAGCAGAAGGAAGAGAAATGAAATACTCATTTTATTGAGTTGGCCC 719
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   49766 AGAAGTATCTACATAAGCAGAAGGAAGAGAAATGAAATACTCATTTTATTGAGTTGGTTT 49707

Qy     720 CCACTGTATGTGGCTGGTATTTATGAAGGTGATGACC-CAGGAAGAAATTGTAAACTATA 778
           ||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db   49706 TCACTGTATGTGGCTGGTATTTATGAAGGTGATGACCTTAGGAAGAAATTGTAAACTATA 49647

Qy     779 AACCACTCCAAATATAAACCCGAGGCAGAAGCAGCATATC-TCCTATGAAGCCTGTATTT 837
           || || |  |||||||||| |||||||||||||||||| |||||||||| || |||||
Db   49646 AATCATTTTAAATATAAACCTGAGGCAGAAGCAGCATATCTTCCTATGAAGTCTATATTT 49587

Qy     838 ACTCAGTGGGAAATAATTTATTAA 861
           |||||||||||||||||||||||
Db   49586 TTTCAGTGGGAAATAATTTATTAA 49563


RESULT   2
HSU57655
LOCUS       HSU57655      626 bp    DNA             PRI       01-SEP-1996
DEFINITION  Human dihydropyrimidine dehydrogenase (DPYD) gene, partial cds.
ACCESSION   U57655
VERSION     U57655.1  GI:1518606
KEYWORDS    .
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 626)
  AUTHORS   Fernandez-Salguero,P.M., Wei,X., Jones,S., Idle,J.R. and
            Gonzalez,F.J.
  TITLE     Lack of dihydropyrimidine dehydrogenase activity and
            thymine-uraciluria caused by a common splice mutation
  JOURNAL   Hum. Mutat. (1996) In press
REFERENCE   2  (bases 1 to 626)
  AUTHORS   Fernandez-Salguero,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-1996) Pedro M. Fernandez-Salguero, National
            Institutes of Health, 9000 Rockville Pike, Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .626
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            179. .343
                     /gene="DPYD"
     CDS             <179. .>343
                     /gene="DPYD"
                     /function="uracil and thymine metabolism"
                     /codon_start=1
                     /evidence=experimental
                     /product="dihydropyrimidine dehydrogenase"
                     /protein_id="AAB07049.1"
                     /db_xref="GI:1518607"
                     /translation="DIVTNVSPRIIRGTTSGPMYGPGQSSFLNIELISEKTAAYWCQS
                     VTELKADFPDN"
BASE COUNT      221 a    100 c    117 g    188 t
ORIGIN
```

```
  Query Match             70.5%;  Score 607;  DB 97;  Length 626;
  Best Local Similarity   99.0%;  Pred. No. 6.8e-109;
  Matches  621;  Conservative    0;  Mismatches    5;  Indels    1;  Gaps    1;

Qy      90 TTTACCTTTTTATTTGCAAGTAGTTTATGTTCAATTCTAATTTAATGTATATTAAAAATT 149
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TTTACCTTTTTATTTGCAAGTAGTTTATGTTCAATTCTAATTTAATGTATATTAAAAATT 60

Qy     150 CCTCTGCAAATATGTGAGGAGGGACCTCATAAAATATTGTCATATGGAAATGAGCAGATA 209
           |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db      61 CCTCTGCAAATATGTGAGGAGGGACCTCAT-AAATATTGTCATATGGAAATGAGCAGATA 119

Qy     210 ATAAAGATTATAGCTTTTCTTTGTCAAAAGGAGACTCAATATCTTTACTCTTTCATGAGG 269
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     120 ATAAAGATTATAGCTTTTCTTTGTCAAAAGGAGACTCAATATCTTTACTCTTTCATCAGG 179

Qy     270 ACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAACCACCTCTGGCCCCATGTATG 329
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     180 ACATTGTGACAAATGTTTCCCCCAGAATCATCCGGGGAACCACCTCTGGCCCCATGTATG 239

Qy     330 GCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCAGTGAGAAAACGGCTGCATATT 389
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     240 GCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCAGTGAGAAAACGGCTGCATATT 299

Qy     390 GGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAGACAACaTAAGTGTGATAAAAA 449
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     300 GGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAGACAACGTAAGTGTGATAAAAA 359

Qy     450 TCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTATTTAAACATCCAATTCATAGG 509
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     360 TCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTATTTAAACATCCAATTCATAGG 419

Qy     510 CTTATAAATATTAATGTGTATATTTTATCAACGAATCTGCCAGTTGCTTTGCTGATGCAT 569
           |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db     420 CTTATAAATATTAATGTGTATATTTTATCAAAGAATCTGCCAGTTGCTTTGCTGATGCAT 479

Qy     570 AGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAAAAACCCACACAATGTGAAGCT 629
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     480 AGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAAAAACCCACACAATGTGAAGCT 539
```

```
Qy     630 CTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATCTACATAAGCAGAAGGAAGAGA 689
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     540 CTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATCTACATAAGTAGAAGGAAGAGA 599

Qy     690 AATGAAATACTCATTTTATTGAGTTGG 716
           ||||||||||||||||||||||||||
Db     600 AATGAAATACTCATTTTATTGAGTTGG 626
```

```
RESULT   3
HSDYPDGEN
LOCUS       HSDYPDGEN     854 bp    DNA             PRI       17-FEB-1997
DEFINITION  H.sapiens DYPD gene, partial sequence.
ACCESSION   X95670
VERSION     X95670.1  GI:1246752
KEYWORDS    dihydropyrimidine dehydrogenase; DYPD gene; exon X; mutation.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 854)
  AUTHORS   Vreken,P., Van Kuilenburg,A.B., Meinsma,R., Smit,G.P., Bakker,H.D.,
            De Abreu,R.A. and van Gennip,A.H.
  TITLE     A point mutation in an invariant splice donor site leads to exon
            skipping in two unrelated Dutch patients with dihydropyrimidine
            dehydrogenase deficiency
  JOURNAL   J. Inherit. Metab. Dis. 19 (5), 645-654 (1996)
  MEDLINE   97047101
REFERENCE   2  (bases 1 to 854)
  AUTHORS   Vreken,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-1996) P. Vreken, Academic Med.Center, Univ.of
            Amsterdam, Dep of Pediatrics and Clinical Chemistry, F0-224, P.O.
            Box 22700, NL-1100 DE Amsterdam, NETHERLANDS
COMMENT     Related sequences:-
            Meinsma et al, DNA Cell Biol. 14:1-6 (1995), U20938, and Yokota H.
            et al, J. Biol. Chem. 269:23192-23196 (1994).
FEATURES             Location/Qualifiers
     source          1. .854
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /tissue_type="liver"
                     /map="p22"
     intron          <1. .568
     gene            569. .734
                     /gene="DYPD"
     CDS             <569. .>734
                     /gene="DYPD"
                     /EC_number="1.3.1.2"
                     /note="exon X"
                     /codon_start=2
                     /product="dihydropyrimidine dehydrogenase (NADP+)"
                     /protein_id="CAA64973.1"
                     /db_xref="GI:1246753"
                     /db_xref="SWISS-PROT:Q12882"
                     /translation="DIVTNVSPRIIRGTTSGPMYGPGQSSFLNIELISEKTAAYWCQS
                     VTELKADFPDN"
     exon            569. .734
                     /gene="DYPD"
     variation       725
                     /gene="DYPD"
                     /replace="c"
     intron          735. .>854
     variation       735
                     /note="changes the splice donor sequence and leads to exon
                     skipping"
                     /replace="a"
BASE COUNT      301 a    126 c    130 g    297 t
ORIGIN
```

```
  Query Match           56.5%;  Score 486.6;  DB 93;  Length 854;
  Best Local Similarity  96.0%;  Pred. No. 2e-85;
  Matches  531;  Conservative    0;  Mismatches   19;  Indels    3;  Gaps    3;

Qy       1 TGTTAATGAAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGT 60
           |||||||||| |||||||||||| ||||| |||| |||||||||||  ||||||||||
Db     304 TGTTAATGAATATAAATATTTTTATTTTTGCGCTATTCTAAACCTAGAATTACAAGAAGT 363

Qy      61 AATTTATCTGGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTT 120
           ||||||  ||  ||||||||||||||||||||||||| ||||||||||||||||||||||
Db     364 AATTTAGCTAAAGCTAACAAATACTTTATTTTACC-TTTTATTTGCAAGTAGTTTATGTT 422

Qy     121 CAATTCTAATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATA 180
           |||||||||||||||||||||||||||||||||||| ||||||| ||||||||||||
Db     423 CAATTCTAATTTAATGTATATTAAAAATTCCTCTGCAAAAATGTGAGAAGGGACCTCATA 482

Qy     181 AAATATTGTCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGG 240
           |||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     483 AAATA-TGTCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGG 541

Qy     241 AGACTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCAT 300
           |||||||||||||||||||||||| |||||||||||||||||||||||| ||||||
Db     542 AGACTCAATATCTTTACTCTTTCATCAGGACATTGTGACAAATGTTTCCCCCAGAATCAT 601

Qy     301 CCGGGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     602 CCGGGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGA 661

Qy     361 GCTCATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     662 GCTCATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGA 721

Qy     421 CTTCCCAGACAACaTAAGTGTGA-TAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGT 479
           ||| |||||||||| |||||||| | || ||||||||||||||||||||||||||||||
Db     722 CTTTCCAGACAACGTAAGTGTGATTTAACATCTAAAACAAGAGAATTGGCATAAGTTGGT 781

Qy     480 GAATGTTTATTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCA 539
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     782 GAATGTTTATTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATTA 841

Qy     540 ACGAATCTGCCAG 552
           | ||||||||||||
Db     842 AAGAATCTGCCAG 854


RESULT   4
HSU09178
LOCUS       HSU09178     3951 bp    mRNA            PRI       28-DEC-1994
DEFINITION  Human dihydropyrimidine dehydrogenase mRNA, complete cds.
ACCESSION   U09178
VERSION     U09178.1  GI:558304
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3951)
  AUTHORS   Yokota,H., Fernandez-Salguero,P., Furuya,H., Lin,K., McBride,O.W.,
            Podschun,B., Schnackerz,K.D. and Gonzalez,F.J.
  TITLE     cDNA cloning and chromosome mapping of human dihydropyrimidine
            dehydrogenase, an enzyme associated with 5-fluorouracil toxicity
            and congenital thymine uraciluria
  JOURNAL   J. Biol. Chem. 269 (37), 23192-23196 (1994)
  MEDLINE   94365020
REFERENCE   2  (sites)
  AUTHORS   Eggink,G., Engel,H., Vriend,G., Terpstra,P. and Witholt,B.
  TITLE     Rubredoxin reductase of Pseudomonas oleovorans. Structural
            relationship to other flavoprotein oxidoreductases based on one NAD
            and two FAD fingerprints
  JOURNAL   J. Mol. Biol. 212 (1), 135-142 (1990)
  MEDLINE   90204534
REFERENCE   3  (sites)
  AUTHORS   Porter,D.J., Chestnut,W.G., Merrill,B.M. and Spector,T.
  TITLE     Mechanism-based inactivation of dihydropyrimidine dehydrogenase by
            5-ethynyluracil
  JOURNAL   J. Biol. Chem. 267 (8), 5236-5242 (1992)
  MEDLINE   92184771
REFERENCE   4  (sites)
  AUTHORS   Dupuis,A., Skehel,J.M. and Walker,J.E.
  TITLE     A homologue of a nuclear-coded iron-sulfur protein subunit of
            bovine mitochondrial complex I is encoded in chloroplast genomes
  JOURNAL   Biochemistry 30 (11), 2954-2960 (1991)
  MEDLINE   91175743
REFERENCE   5  (sites)
  AUTHORS   Wierenga,R.K., De Maeyer,M.C.H. and Hol,W.G.J.
  TITLE     Interaction of pyrophosphatase moieties with alfa-helixes in
            dinucleotide binding proteins
  JOURNAL   Biochemistry 24, 1346-1357 (1985)
REFERENCE   6  (bases 1 to 3951)
  AUTHORS   Gonzalez,F.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-APR-1994) Frank J. Gonzalez, National Cancer
            Institute, National Institutes of Health, 9000 Rockville Pike,
            Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .3951
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="liver"
     CDS             82. .3159
                     /codon_start=1
                     /product="dihydropyrimidine dehydrogenase"
                     /protein_id="AAA57474.1"
                     /db_xref="GI:558305"
                     /translation="MAPVLSKDSADIESILALNPRTQTHATLCSTSAKKLDKKHWKRN
                     PDKNCFNCEKLENNFDDIKHTTLGERGALREAMRCLKCADAPCQKSCPTNLDIKSFIT
                     SIANKNYYGAAKMIFSDNPLGLTCGMVCPTSDLCVGGCNLYATEEGPINIGGLQQFAT
                     EVFKAMSIPQIRNPSLPPPEKMSEAYSAKIALFGAGPASISCASFLARLGYSDITIFE
                     KQEYVGGLSTSEIPQFRLPYDVVNFEIELMKDLGVKIICGKSLSVNEMTLSTLKEKGY
                     KAAFIGIGLPEPNKDAIFQGLTQDQGFYTSKDFLPLVAKGSKAGMCACHSPLPSIRGV
                     VIVLGAGDTAFDCATSALRCGARRVFIVFRKGFVNIRAVPEEMELAKEEKCEFLPFLS
                     PRKVIVKGGRIVAMQFVRTEQDETGKWNEDEDQMVHLKADVVISAFGSVLSDPKVKEA
                     LSPIKFNRWGLPEVDPETMQTSEAWVFAGGDVVGLANTTVESVNDGKQASWYIHKYVQ
                     SQYGASVSAKPELPLFYTPIDLVDISVEMAGLKFINPFGLASATPATSTSMIRRAFEA
                     GWGFALTKTFSLDKDIVTNVSPRIIRGTTSGPMYGPGQSSFLNIELISEKTAAYWCQS
                     VTELKADFPDNIVIASIMCSYNKNDWTELAKKSEDSGADALELNLSCPHGMGERGMGL
                     ACGQDPELVRNICRWVRQAVQIPFFAKLTPNVTDIVSIARAAKEGGANGVTATNTVSG
                     LMGLKSDGTPWPAVGIAKRTTYGGVSGTAIRPIALRAVTSIARALPGFPILATGGIDS
                     AESGLQFLHSGASVLQVCSAIQNQDFTVIEDYCTGLKALLYLKSIEELQDWDGQSPAT
                     VSHQKGKPVPRIAELMDKKLPSFGPYLEQRKKIIAENKIRLKEQNVAFSPLKRSCFIP
                     KRPIPTIKDVIGKALQYLGTFGELSNVEQVVAMIDEEMCINCGKCYMTCNDSGYQAIQ
                     FDPETHLPTITDTCTGCTLCLSVCPIVDCIKMVSRTTPYEPKRGVPLSVNPVC"
     misc_feature    1084. .1134
                     /citation=[5]
                     /function="catalytic cofactor NADPH/NADP binding site"
     misc_feature    1492. .1524
                     /citation=[2]
                     /function="electron transfer center, FAD binding site"
     misc_feature    2062. .2175
                     /citation=[3]
                     /function="uracil (substrate) binding site"
     misc_feature    2938. .2973
                     /standard_name="iron-sulfur center"
                     /citation=[4]
                     /function="catalytic cofactor [4Fe-4S] binding site"
BASE COUNT     1153 a    785 c    896 g   1117 t
ORIGIN


  Query Match           20.1%;  Score 172.8;  DB 97;  Length 3951;
  Best Local Similarity  93.8%;  Pred. No. 2.8e-24;
  Matches  180;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy     244 CTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCG 303
           | | | | ||||     || || |||||||||||||||||||||||||| |||||||||
Db    1797 CACCAAAACTTTCTCTCTTGATAAGGACATTGTGACAAATGTTTCCCCCAGAATCATCCG 1856
```

```
Qy     304 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1857 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 1916

Qy     364 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 423
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1917 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 1976

Qy     424 CCCAGACAACaT 435
           ||||||||||||
Db    1977 CCCAGACAACAT 1988


RESULT   5
AR027740
LOCUS       AR027740     3957 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 1 from patent US 5856454.
ACCESSION   AR027740
VERSION     AR027740.1  GI:5938560
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3957)
  AUTHORS   Gonzalez,F.J. and Fernandez-Salguero,P.
  TITLE     CDNA for human and pig dihydropyrimidine dehydrogenase
  JOURNAL   Patent: US 5856454-A 1 05-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .3957
                     /organism="unknown"
BASE COUNT     1156 a    787 c    897 g   1117 t
ORIGIN
 .

  Query Match             20.1%;  Score 172.8;  DB 9;  Length 3957;
  Best Local Similarity   93.8%;  Pred. No. 2.8e-24;
  Matches  180;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;
 .
Qy     244 CTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCG 303
           | | | | ||||     || || ||||||||||||||||||||||||||| |||||||||
Db    1803 CACCAAAACTTTCTCTCTTGATAAGGACATTGTGACAAATGTTTCCCCCAGAATCATCCG 1862

Qy     304 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1863 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 1922

Qy     364 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 423
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1923 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 1982

Qy     424 CCCAGACAACaT 435
           ||||||||||||
Db    1983 CCCAGACAACAT 1994


RESULT   6
AB003063
LOCUS       AB003063     3078 bp    mRNA            PRI       21-JAN-2000
DEFINITION  Homo sapiens mRNA for dihydropyrimidine dehydrogenase, complete
            cds.
ACCESSION   AB003063
VERSION     AB003063.1  GI:6729337
KEYWORDS    dihydropyrimidine dehydrogenase.
SOURCE      Homo sapiens male lymphocyte cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ogura,K., Nishiyama,T., Takubo,H., Kato,A., Okuda,H., Arakawa,K.,
            Fukushima,M., Nagayama,S., Kawaguchi,Y. and Watabe,T.
  TITLE     Suicidal inactivation of human dihydropyrimidine dehydrogenase by
            (E)-5-(2-bromovinyl)uracil derived from the antiviral, sorivudine
 .
  JOURNAL   Cancer Lett. 122 (1-2), 107-113 (1998)
  MEDLINE   98124145
  REMARK    Erratum:[[published erratum appears in Cancer Lett 1998 Jun
            19;128(2):229]]
REFERENCE   2  (bases 1 to 3078)
  AUTHORS   Ogura,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-1997) to the DDBJ/EMBL/GenBank databases.
            Kenichiro Ogura, Tokyo University of Pharmacy and Life Science,
            Department of Drug Metabolism and Molecular Toxicology; 1432-1
            Horinouchi, Hachioji, Tokyo 192-0392, Japan
            (E-mail:ogurak@ps.toyaku.ac.jp, Tel:+81-426-76-4518,
            Fax:+81-426-76-4517)
FEATURES             Location/Qualifiers
     source          1. .3078
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="lymphocyte"
                     /sex="male"
     CDS             1. .3078
                     /codon_start=1
                     /product="dihydropyrimidine dehydrogenase"
                     /protein_id="BAA89789.1"
                     /db_xref="GI:6729338"
                     /translation="MAPVLSKDSADIESILALNPRTQTHATLCSTSAKKLDKKHWKRN
                     PDKNCFNCEKLENNFDDIKHTTLGERGALREAMRCLKCADAPCQKSCPTNLDIKSFIT
                     SIANKNYYGAAKMIFSDNPLGLTCGMVCPTSDLCVGGCNLYATEEGPINIGGLQQFAT
                     EVFKAMSIPQIRNPSLPPPEKMSEAYSAKIALFGAGPASISCASFLARLGYSDITIFE
                     KQEYVGGLSTSEIPQFRLPYDVVNFEIELMKDLGVKIICGKSLSVNEMTLSTLKEKGY
                     KAAFIGIGLPEPNKDAIFQGLTQDQGFYTSKDFLPLVAKGSKAGMCACHSPLPSIRGV
                     VIVLGAGDTAFDCATSALRCGARRVFIVFRKGFVNIRAVPEEMELAKEEKCEFLPFLS
                     PRKVIVKGGRIVAMQFVRTEQDETGKWNEDEDQMVHLKADVVISAFGSVLSDPKVKEA
                     LSPIKFNRWGLPEVDPETMQTSEAWVFAGGDVVGLANTTVESVNDGKQASWYIHKYVQ
                     SQYGASVSAKPELPLFYTPIDLVDISVEMAGLKFINPFGLASATPATSTSMIRRAFEA
                     GWGFALTKTFSLDKDIVTNVSPRIIRGTTSGPMYGPGQSSFLNIELISEKTAAYWCQS
                     VTELKADFPDNIVIASIMCSYNKNDWTELAKKSEDSGADALELNLSCPHGMGERGMGL
                     ACGQDPELVRNICRWVRQAVQIPFFAKLTPNVTDIVSIARAAKEGGANGVTATNTVSG
                     LMGLKSDGTPWPAVGIAKRTTYGGVSGTAIRPIALRAVTSIARALPGFPILATGGIDS
                     AESGLQFLHSGASVLQVCSAIQNQDFTVIEDYCTGLKALLYLKSIEELQDWDGQSPAT
                     VSHQKGKPVPRIAELMDKKLPSFGPYLEQRKKIIAENKIRLKEQNVAFSPLKRNCFIP
                     KRPIPTIKDVIGKALQYLGTFGELSNVEQVVAMIDEEMCINCGKCYMTCNDSGYQAIQ
                     FDPETHLPTITDTCTGCTLCLSVCPIVDCIKMVSRTTPYEPKRGVPLSVNPVC"
BASE COUNT      875 a    643 c    734 g    826 t
ORIGIN


  Query Match             19.9%;  Score 171.2;  DB 85;  Length 3078;
  Best Local Similarity   93.2%;  Pred. No. 5.8e-24;
  Matches  179;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy     244 CTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCG 303
           | | | | ||||     || || ||||||||||||||||||||||||||| |||||||||
Db    1716 CACCAAAACTTTCTCTCTTGATAAGGACATTGTGACAAATGTTTCCCCCAGAATCATCCG 1775

Qy     304 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1776 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 1835

Qy     364 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 423
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1836 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 1895

Qy     424 CCCAGACAACaT 435
           ||||||||||||
Db    1896 TCCAGACAACAT 1907


RESULT   7
HSU20938
LOCUS       HSU20938     4409 bp    mRNA            PRI       04-APR-1997
DEFINITION  Human lymphocyte dihydropyrimidine dehydrogenase mRNA, complete
            cds.
ACCESSION   U20938
```

```
VERSION     U20938.1  GI:1926407
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4409)
  AUTHORS   Diasio,R.B., Beavers,T.L. and Carpenter,J.T.
  TITLE     Familial deficiency of dihydropyrimidine dehydrogenase. Biochemical
            basis for familial pyrimidinemia and severe 5-fluorouracil-induced
            toxicity
  JOURNAL   J. Clin. Invest. 81 (1), 47-51 (1988)
  MEDLINE   88087908
REFERENCE   2  (bases 1 to 4409)
  AUTHORS   Lu,Z.H., Zhang,R. and Diasio,R.B.
  TITLE     Purification and characterization of dihydropyrimidine
            dehydrogenase from human liver
  JOURNAL   J. Biol. Chem. 267 (24), 17102-17109 (1992)
  MEDLINE   92381021
REFERENCE   3  (bases 1 to 4409)
  AUTHORS   Johnson,M.R., Albin,N., Shahinian,H. and Diasio,R.B.
  TITLE     Identification of a frameshift in the DNA coding for
            dihydropyrimidine dehydrogenase (DPD) in a DPD deficient patient
            exhibiting 5-fluorouracil toxicity
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 4409)
  AUTHORS   Johnson,M.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-1995) Pharmacology, University of Alabama at
            Birmingham, 1670 University Blvd., Birmingham, AL 35294-0019, USA
COMMENT     On Apr 4, 1997 this sequence version replaced gi:693911.
FEATURES             Location/Qualifiers
     source          1. .4409
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="lymphocyte"
     CDS             102. .3179
                     /EC_number="1.2.1.2"
                     /note="DPD; dihydrouracil dehydrogenase"
                     /codon_start=1
                     /product="dihydropyrimidine dehydrogenase"
                     /protein_id="AAB51366.1"
                     /db_xref="GI:693912"
                     /translation="MAPVLSKDSADIESILALNPRTQTHATLCSTSAKKLDKKHWKRN
                     PDKNCFNCEKLENNFDDIKHTTLGERGALREAMRCLKCADAPCQKSCPTNLDIKSFIT
                     SIANKNYYGAAKMIFSDNPLGLTCGMVCPTSDLCVGGCNLYATEEGPINIGGLQQFAT
                     EVFKAMSIPQIRNPSLPPPEKMSEAYSAKIALFGAGPASISCASFLARLGYSDITIFE
                     KQEYVGGLSTSEIPQFRLPYDVVNFEIELMKDLGVKIICGKSLSVNEMTLSTLKEKGY
                     KAAFIGIGLPEPNKDAIFQGLTQDQGFYTSKDFLPLVAKGSKAGMCACHSPLPSIRGV
                     VIVLGAGDTAFDCATSALRCGARRVFIVFRKGFVNIRAVPEEMELAKEEKCEFLPFLS
                     PRKVIVKGGRIVAMQFVRTEQDETGKWNEDEDQMVHLKADVVISAFGSVLSDPKVKEA
                     LSPIKFNRWGLPEVDPETMQTSEAWVFAGGDVVGLANTTVESVNDGKQASWYIHKYVQ
                     SQYGASVSAKPELPLFYTPIDLVDISVEMAGLKFINPFGLASATPATSTSMIRRAFEA
                     GWGFALTKTFSLDKDIVTNVSPRIIRGTTSGPMYGPGQSSFLNIELISEKTAAYWCQS
                     VTELKADFPDNIVIASIMCSYNKNDWTELAKKSEDSGADALELNLSCPHGMGERGMGL
                     ACGQDPELVRNICRWVRQAVQIPFFAKLTPNVTDIVSIARAAKEGGANGVTATNTVSG
                     LMGLKSDGTPWPAVGIAKRTTYGGVSGTAIRPIALRAVTSIARALPGFPILATGGIDS
                     AESGLQFLHSGASVLQVCSAIQNQDFTVIEDYCTGLKALLYLKSIEELQDWDGQSPAT
                     VSHQKGKPVPRIAELMDKKLPSFGPYLEQRKKIIAENKIRLKEQNVAFSPLKRNCFIP
                     KRPIPTIKDVIGKALQYLGTFGELSNVEQVVAMIDEEMCINCGKCYMTCNDSGYQAIQ
                     FDPETHLPTITDTCTGCTLCLSVCPIVDCIKMVSRTTPYEPKRGVPLSVNPVC"
BASE COUNT     1317 a    845 c    949 g   1298 t
ORIGIN
```

```
  Query Match           19.9%;  Score 171.2;  DB 97;  Length 4409;
  Best Local Similarity  93.2%;  Pred. No. 5.8e-24;
  Matches  179;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy     244 CTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCG 303
           | | | | ||||     || || ||||||||||||||||||||||||||| |||||||||
Db    1817 CACCAAAACTTTCTCTCTTGATAAGGACATTGTGACAAATGTTTCCCCCAGAATCATCCG 1876

Qy     304 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1877 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 1936

Qy     364 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 423
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1937 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 1996

Qy     424 CCCAGACAACaT 435
           ||||||||||||
Db    1997 TCCAGACAACAT 2008
```

```
RESULT   8
BTU20981
LOCUS       BTU20981     4421 bp    mRNA            MAM       17-JAN-1997
DEFINITION  Bos taurus liver dihydropyrimidine dehydrogenase mRNA, complete
            cds.
ACCESSION   U20981
VERSION     U20981.1  GI:677950
KEYWORDS    .
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 4421)
  AUTHORS   Lu,Z., Zhang,R. and Diasio,R.B.
  TITLE     Dihydropyrimidine dehydrogenase activity in human peripheral blood
            mononuclear cells and liver: population characteristics, newly
            identified deficient patients, and clinical implication in
            5-fluorouracil chemotherapy
  JOURNAL   Cancer Res. 53 (22), 5433-5438 (1993)
  MEDLINE   94036837
REFERENCE   2  (bases 1 to 4421)
  AUTHORS   Albin,N., Johnson,M.R. and Diasio,R.B.
  TITLE     cDNA cloning of bovine liver dihydropyrimidine dehydrogenase
  JOURNAL   DNA Seq. 6 (4), 243-250 (1996)
  MEDLINE   97069948
REFERENCE   3  (bases 1 to 4421)
  AUTHORS   Johnson,M.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-FEB-1995) Pharmacology/Toxicology, University of
            Alabama at Birmingham, 1670 University Blvd., Birmingham, AL
            35294-0019, USA
FEATURES             Location/Qualifiers
     source          1. .4421
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /tissue_type="liver"
     CDS             75. .3152
                     /EC_number="1.3.1.2"
                     /note="dihydrothymine dehydrogenase; DPD"
                     /codon_start=1
                     /product="dihydropyrimidine dehydrogenase"
                     /protein_id="AAB40985.1"
                     /db_xref="GI:677951"
                     /translation="MAPVLSKDVADIESILALNPRTQSRATLRSTLAKKLDKKHWKRN
                     PDKNCFNCEKLENNFDDIKHTTLGERGALREAMRCLKCADAPCQKSCPTNLDIKSFIT
                     SISNKNYYGAAKMIFSDNPLGLTCGMVCPTSDLCVGGCNLYATEEGPINIGGLQQYAT
                     EVFKAMNIPQIRNPSLPPPEKMPEAYSAKIALLGAGPASISCASFLARLGYNDITIFE
                     KQEYVGGISTSEIPQFRLPYDVVNFEIELMKDLGVKIICGKSLSVNDITLSTLKEEGY
                     KAAFIGIGLPEPKKDHIFQGLTQDQGFYTSKDFLPLVAKSSKAGMCACHSPLLSIRGT
                     VIVLGAGDTAFDCATSALRCGARRVFIVFRKGFVNIRAVPEEVELAREEKCEFLPFLS
                     PRKVIVKGGRIVAMQFVRTEQDETGKWNEDGDQIACLKADVVISAFGSVLSDPKVKEA
                     LSPIKFNRWDLPEVDPETMQTSEPWVFAGGDVVGIANTTVEAVNDGKQASWYIHRYIQ
                     SQYGASVSAKPELPLFYTPIDLVDISVEMAALKFTNPFGLASATPTTSSSMIRRAFEA
                     GWAFALTKTFSLDKDIVTNVSPRIIRGTTSGPMYGPGQSSFLNIELISEKTAAYWCQS
                     VTELKADFPDNIVIASIMCSYNRNDWMELSRKAEASGADALELNLSCPHGMGERGMGL
                     ACGQDPELVRNICRWVRQAVRIPFFAKLTPNVTDIVSIARAAKEGGANGVTATNTVSG
                     LMGLKADGTPWPAVGREKRTTYGGVSGTAIRPIALRAVTTIARALPEFPILATGGIDS
                     AESGLQFLHGGASVLQVCSAIQNQDFTIIQDYCTGLKALLYLKSIEELQDWDGQSPAT
                     KSHQKGKPVPCIAELVGKKLPSFGPYLEKCKKIIAEEKLRLKKENVTVLPLERNHFIP
```

```
                    KKPIPSVKDVIGKALQYLGTYGELNNTEQVVAVIDEEMCINCGKCYMTCNDSGYQAIQ
                    FDPETHLPTVTDTCTGCTLCLSVCPIIDCIKMVSRTTPYEPKRGLPLAVNPVS"
BASE COUNT     1315 a    874 c    958 g   1274 t
ORIGIN


  Query Match             18.0%;  Score 154.8;  DB 7;  Length 4421;
  Best Local Similarity   88.4%;  Pred. No. 9.1e-21;
  Matches  168;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy     246 CAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGG 305
           | | | ||||     || || |||| || ||||||||||||||| |||| ||||||||||
Db    1792 CCAAAACTTTCTCTCTTGATAAGGATATAGTGACAAATGTTTCACCCAGAATCATCCGGG 1851

Qy     306 GAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCA 365
           | |||||||||||||||||||||||||||||||||||||||| || ||||||||||||||
Db    1852 GGACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCTTTCCTGAATATTGAGCTCA 1911

Qy     366 TCAGTGAGAAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCC 425
           ||||||| |||||||||||||||||||||||||||||||||||||||||| || ||||| |
Db    1912 TCAGTGAAAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAAGCCGACTTTC 1971

Qy     426 CAGACAACaT 435
           ||||||| ||
Db    1972 CAGACAATAT 1981


RESULT   9
SSU09179
LOCUS       SSU09179     4441 bp    mRNA            MAM       28-DEC-1994
DEFINITION  Sus scrofa dihydropyrimidine dehydrogenase mRNA, complete cds.
ACCESSION   U09179
VERSION     U09179.1  GI:558306
KEYWORDS    .
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 4441)
  AUTHORS   Yokota,H., Fernandez-Salguero,P., Furuya,H., Lin,K., McBride,O.W.,
            Podschun,B., Schnackerz,K.D. and Gonzalez,F.J.
  TITLE     cDNA cloning and chromosome mapping of human dihydropyrimidine
            dehydrogenase, an enzyme associated with 5-fluorouracil toxicity
            and congenital thymine uraciluria
  JOURNAL   J. Biol. Chem. 269 (37), 23192-23196 (1994)
  MEDLINE   94365020
REFERENCE   2  (sites)
  AUTHORS   Eggink,G., Engel,H., Vriend,G., Terpstra,P. and Witholt,B.
  TITLE     Rubredoxin reductase of Pseudomonas oleovorans. Structural
            relationship to other flavoprotein oxidoreductases based on one NAD
            and two FAD fingerprints
  JOURNAL   J. Mol. Biol. 212 (1), 135-142 (1990)
  MEDLINE   90204534
REFERENCE   3  (sites)
  AUTHORS   Porter,D.J., Chestnut,W.G., Merrill,B.M. and Spector,T.
  TITLE     Mechanism-based inactivation of dihydropyrimidine dehydrogenase by
            5-ethynyluracil
  JOURNAL   J. Biol. Chem. 267 (8), 5236-5242 (1992)
  MEDLINE   92184771
REFERENCE   4  (sites)
  AUTHORS   Dupuis,A., Skehel,J.M. and Walker,J.E.
  TITLE     A homologue of a nuclear-coded iron-sulfur protein subunit of
            bovine mitochondrial complex I is encoded in chloroplast genomes
  JOURNAL   Biochemistry 30 (11), 2954-2960 (1991)
  MEDLINE   91175743
REFERENCE   5  (sites)
  AUTHORS   Wierenga,R.K., De Maeyer,M.C.H. and Hol,W.G.J.
  TITLE     Interaction of pyrophosphatase moieties with alfa-helixes in
            dinucleotide binding proteins
  JOURNAL   Biochemistry 24, 1346-1357 (1985)
REFERENCE   6  (bases 1 to 4441)
  AUTHORS   Gonzalez,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-APR-1994) Frank J. Gonzalez, National Institutes of
            Health, National Cancer Institute, 9000 Rockville Pike, Bethesda,
            MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .4441
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /tissue_type="liver"
     CDS             82. .3159
                     /codon_start=1
                     /product="dihydropyrimidine dehydrogenase"
                     /protein_id="AAA57475.1"
                     /db_xref="GI:558307"
                     /translation="MAPVLSKDVADIESILALNPRTQSHAALHSTLAKKLDKKHWKRN
                     PDKNCFHCEKLENNFGDIKHTTLGERGALREAMRCLKCADAPCQKSCPTHLDIKSFIT
                     SISNKNYYGAAKMIFSDNPLGLTCGMVCPTSDLCVGGCNLYATEEGSINIGGLQQFAS
                     EVFKAMNIPQIRNPCLPSQEKMPEAYSAKIALLGAGPASISCASFLARLGYSDITIFE
                     KQEYVGGLSTSEIPQFRLPYDVVNFEIELMKDLGVKIICGKSLSENEITLNTLKEEGY
                     KAAFIGIGLPEPKTDDIFQGLTQDQGFYTSKDFLPLVAKSSKAGMCACHSPLPSIRGA
                     VIVLGAGDTAFDCATSALRCGARRVFLVFRKGFVNIRAVPEEVELAKEEKCEFLPFLS
                     PRKVIVKGGRIVAVQFVRTEQDETGKWNEDEDQIVHLKADVVISAFGSVLRDPKVKEA
                     LSPIKFNRWDLPEVDPETMQTSEPWVFAGGDIVGMANTTVESVNDGKQASWYIHKYIQ
                     AQYGASVSAKPELPLFYTPVDLVDISVEMAGLKFINPFGLASAAPTTSSSMIRRAFEA
                     GWGFALTKTFSLDKDIVTNVSPRIVRGTTSGPMYGPGQSSFLNIELISEKTAAYWCQS
                     VTELKADFPDNIVIASIMCSYNKNDWMELSRKAEASGADALELNLSCPHGMGERGMGL
                     ACGQDPELVRNICRWVRQAVQIPFFAKLTPNVTDIVSIARAAKEGGADGVTATNTVSG
                     LMGLKADGTPWPAVGAGKRTTYGGVSGTAIRPIALRAVTTIARALPGFPILATGGIDS
                     AESGLQFLHSGASVLQVCSAVQNQDFTVIQDYCTGLKALLYLKSIEELQGWDGQSPGT
                     ESHQKGKPVPRIAELMGKKLPNFGPYLEQRKKIIAEEKMRLKEQNAAFPPLERKPFIP
                     KKPIPAIKDVIGKALQYLGTFGELSNIEQVVAVIDEEMCINCGKCYMTCNDSGYQAIQ
                     FDPETHLPTVTDTCTGCTLCLSVCPIIDCIRMVSRTTPYEPKRGLPLAVNPVC"
     misc_feature    1084. .1134
                     /citation=[5]
                     /function="catalytic cofactor NADPH/NADP binding site"
     misc_feature    1492. .1524
                     /citation=[2]
                     /function="electron transfer center, FAD binding site"
     misc_feature    2062. .2175
                     /citation=[3]
                     /function="uracil (substrate) binding site"
     misc_feature    2938. .2973
                     /standard_name="iron-sulfur center"
                     /citation=[4]
                     /function="catalytic cofactor [4Fe-4S] binding site"
BASE COUNT     1281 a    928 c    998 g   1234 t
ORIGIN


  Query Match             17.6%;  Score 151.6;  DB 7;  Length 4441;
  Best Local Similarity   87.4%;  Pred. No. 3.8e-20;
  Matches  166;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy     246 CAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGG 305
           | | | ||||     || || ||||||| |||||||||| || |||| |||| ||||||
Db    1799 CCAAAACTTTCTCTCTTGATAAGGACATAGTGACAAATGTCTCACCCAGAATCGTCCGGG 1858

Qy     306 GAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCA 365
           | || ||||||||||||||||| ||||||||||||||||||||| ||||||||||||||||
Db    1859 GGACTACCTCTGGCCCCATGTACGGCCCTGGACAAAGCTCCTTCCTGAATATTGAGCTCA 1918

Qy     366 TCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCC 425
           ||||||| ||||| |||||||||||||||||||||||||||||||||||| |||||||| |
Db    1919 TCAGTGAAAAAACAGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAAGCTGACTTTC 1978

Qy     426 CAGACAACaT 435
           ||||||| ||
Db    1979 CAGACAATAT 1988


RESULT  10
AR027741
LOCUS       AR027741     4447 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 3 from patent US 5856454.
```

```
ACCESSION   AR027741
VERSION     AR027741.1  GI:5938561
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4447)
  AUTHORS   Gonzalez,F.J. and Fernandez-Salguero,P.
  TITLE     CDNA for human and pig dihydropyrimidine dehydrogenase
  JOURNAL   Patent: US 5856454-A 3 05-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .4447
                     /organism="unknown"
BASE COUNT     1283 a    930 c   1000 g   1234 t
ORIGIN
```

```
  Query Match             17.6%;  Score 151.6;  DB 9;  Length 4447;
  Best Local Similarity   87.4%;  Pred. No. 3.8e-20;
  Matches  166;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy     246 CAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGG 305
           | | | ||||     || || ||||||| |||||||||| || |||| |||| ||||||
Db    1805 CCAAAACTTTCTCTCTTGATAAGGACATAGTGACAAATGTCTCACCCAGAATCGTCCGGG 1864

Qy     306 GAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCA 365
           | || ||||||||||||||||| |||||||||||||||||| ||||||||||||||||
Db    1865 GGACTACCTCTGGCCCCATGTACGGCCCTGGACAAAGCTCCTTCCTGAATATTGAGCTCA 1924

Qy     366 TCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCC 425
           ||||||| ||||| |||||||||||||||||||||||||||||||||| |||||||| |
Db    1925 TCAGTGAAAAAACAGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAAGCTGACTTTC 1984

Qy     426 CAGACAACaT 435
           ||||||| ||
Db    1985 CAGACAATAT 1994
```

```
RESULT  11
D85035
LOCUS       D85035       4358 bp    mRNA            ROD       19-SEP-1998
DEFINITION  Rattus norvegicus mRNA for dihydropyrimidine dehydrogenase,
            complete cds.
ACCESSION   D85035
VERSION     D85035.1  GI:3628592
KEYWORDS    DPD; dihydropyrimidine dehydrogenase.
SOURCE      Rattus norvegicus (strain:Wistar) 7-d-old male liver cDNA to mRNA,
            clone_lib:Lambda ZAPII cDNA.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 4358)
  AUTHORS   Kimura,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-1996) to the DDBJ/EMBL/GenBank databases.
            Masahiro Kimura, Kobe Gakuin University, Faculty of Nutrition; 518
            Arise, Ikawadani-cho, Nishi-ku, Kobe, Hyogo 651-21, Japan
            (Tel:078-974-1551(ex.3242), Fax:078-974-5689)
REFERENCE   2  (bases 1 to 4358)
  AUTHORS   Kimura,M., Sakata,S.F., Matoba,Y., Matsuda,K., Kontani,Y.,
            Kaneko,M. and Tamaki,N.
  TITLE     Cloning of rat dihydropyrimidine dehydrogenase and correlation of
            its mRNA increase in the rat liver with age
  JOURNAL   J. Nutr. Sci. Vitaminol. 44 (4), 537-546 (1998)
  MEDLINE   99037149
FEATURES             Location/Qualifiers
     source          1. .4358
                     /organism="Rattus norvegicus"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /clone_lib="Lambda ZAPII cDNA"
                     /dev_stage="7-d-old"
                     /sex="male"
                     /tissue_type="liver"
     gene            59. .3136
                     /gene="DPD"
     CDS             59. .3136
                     /gene="DPD"
                     /EC_number="1.3.1.2"
                     /codon_start=1
                     /product="dihydropyrimidine dehydrogenase"
                     /protein_id="BAA33218.1"
                     /db_xref="GI:3628593"
                     /translation="MAGVLSRDAPDIESILALNPRIQAHATLRSTMAKKLDKKHWKRN
                     TDKNCFICEKLENNFDDIKHTTLGERGALREAVRCLKCADAPCQKSCPTSLDIKSFIT
                     SIANKNYYGAAKLIFSDNPLGLTCGMVCPTSDLCVGGCNLHATEEGPINIGGLQQFAT
                     EVFKAMNIPQIRSPLLPPPEHMPEAYSAKIALFGAGPASISCASFLARLGYSDITIFE
                     KQEYVGGLSTSEIPQFRLPYDVVNFEIELMKDLGVKIICGKSISTDEMTLSTLKENGY
                     KAAFIGIGLPEPKKDHIFQGLTQVQGFYTSKDFLPLVAKGSKPGMCACHSPLPSVRGA
                     VIVLGAGDTAFDCATSALRCGARRVFIVFRKGFANIRAVPEEMELAKEEKCEFLPFLS
                     PRKVIVKDGKIVGMQFVRTEQDETGNWVEDEEQIVRLKADVVISPFGSVLDDPKVIEA
                     LSPIKFNRWGLPEVNPETMQTSEPWVFAGGDVVGMANTTVESVNDGKQASWYIHEYIQ
                     AQYGALVPSQPTLPLFYTPVDLVDISVEMAGLRFPNPFGLASATPATSTPMIRRAFEA
                     GWGFALTKTFSLDKDIVTNVSPRIIRGTTSGPLYGPGQSSFLNIELISEKTAAYWCHS
                     VTELKADFPDNILIASIMCSYNKNDWMELSKMAEASGADALELNLSCPHGMGERGMGL
                     ACGQDPELVRNICRWVRQSVRVPFFAKLTPNVTDIVSIARAAKEGGADGVTATNTVSG
                     LMGLKADGSPWPSVGSGKRTTYGGVSGTTIRPIALRAVTAIARALPGFPILATGGIDS
                     AESGLQFLHSGASVLQVCSAIQNQDFTVIEDYCTGLKALLYLKSIEELSDWDGQSPPT
                     MSHQKGKPVPHIAELMGQKLPSFGPYLERRKKILAASKIRENDQNRACSPLQRKHFNS
                     QKPIPAIKDVIGKSLQYLGTFGELNIMEQVVALIDEEMCINCGKCYMTCNDSGYQAIQ
                     FDPETHLPTVSDTCTGCTLCLSVCPIMDCIRMVSRATPYEPKRGLPLAVKPVC"
BASE COUNT     1234 a    930 c   1022 g   1172 t
ORIGIN
```

```
  Query Match             17.0%;  Score 146.8;  DB 94;  Length 4358;
  Best Local Similarity   85.8%;  Pred. No. 3.3e-19;
  Matches  163;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy     246 CAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGG 305
           | | | ||||     || || ||||||| |||||||| || || |||| ||||||||| |
Db    1776 CCAAAACTTTCTCTCTTGATAAGGACATCGTGACAAACGTCTCACCCAGAATCATCCGAG 1835

Qy     306 GAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCA 365
           | ||||| ||||||||| ||||||||||||||||||||||| || || || ||||||||||
Db    1836 GGACCACTTCTGGCCCCTTGTATGGCCCTGGACAAAGCTCTTTCCTCAACATTGAGCTCA 1895

Qy     366 TCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCC 425
           |||||||||||| |||||||||||||||| |||||||| ||||||||||||||||||||
Db    1896 TCAGTGAGAAAACAGCTGCATATTGGTGTCACAGTGTCACCGAACTAAAGGCTGACTTCC 1955

Qy     426 CAGACAACaT 435
           | ||||||||
Db    1956 CGGACAACAT 1965
```

```
RESULT  12
CELC25F6
LOCUS       CELC25F6    43803 bp    DNA             INV       02-NOV-1995
DEFINITION  Caenorhabditis elegans cosmid C25F6.
ACCESSION   U39742
VERSION     U39742.1  GI:1049455
KEYWORDS    .
SOURCE      Caenorhabditis elegans strain=Bristol N2.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 43803)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
```

```
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 43803)
  AUTHORS   Bentley,D.
  TITLE     The sequence of C. elegans cosmid C25F6
  JOURNAL   Unpublished (1995)
REFERENCE   3  (bases 1 to 43803)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-OCT-1995) Robert Waterston
COMMENT               Submitted by:
                      Genome Sequencing Center
                      Department of Genetics, Washington University,
                      St. Louis, MO  63110, USA, and
                      Sanger Centre, Hinxton Hall
                      Cambridge CB10 IRQ, England
                      e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk
                      NEIGHBORING COSMID INFORMATION:
            The 5' cosmid is C41A3, 1000 bp overlap; 3' end lies in a gap
            followed by the cosmid T23F2. Actual start of this cosmid is at
            base position 1 of CELC25F6; actual end is at base position 43803
            of CELC25F6.

            NOTES:
            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES              Location/Qualifiers
     source           1. .43803
                      /organism="Caenorhabditis elegans"
                      /db_xref="taxon:6239"
                      /map="X"
     tRNA             complement(1397. .1495)
                      /note="Leu;  codon recognized: CTA"
     gene             complement(5095. .5656)
                      /gene="C25F6.6"
     CDS              complement(join(5095. .5215,5259. .5385,5443. .5535,
                      5584. .5656))
                      /gene="C25F6.6"
                      /codon_start=1
                      /evidence=not_experimental
                      /protein_id="AAA80431.1"
                      /db_xref="GI:1049456"
                      /translation="MMISRFLSKILIVSCLIGVLGVADASSSMYARRRTASERLMSII
                      RPRMFQNPEFRSSQKIQKSIQLNPNRALPCCKDESGGSICKNLRRTDIKLFTQKCQTE
                      PDFSLVVCCSSCSDAGISYRKRAQKQVFIFTILVI"
     tRNA             complement(6056. .6128)
                      /note="Lys;  codon recognized: AAA"
     gene             16243. .20895
                      /gene="C25F6.4"
     CDS              join(16243. .16401,16449. .16601,17110. .17257,17309. .17421,
                      17576. .17675,17720. .17825,18125. .18202,18252. .18358,
                      18406. .18572,18930. .19243,19293. .19592,20318. .20513,
                      20569. .20718,20773. .20895)
                      /gene="C25F6.4"
                      /note="similar to protein tyrosine kinase"
                      /codon_start=1
                      /evidence=not_experimental
                      /protein_id="AAA80432.1"
                      /db_xref="GI:1049457"
                      /translation="MQNGDIADSQITASSSFDKQSVGPQNARLHSELASGAWCPKPQI
                      NSKSYEFLQVTLNDTFLITSVETQGRYGNGTGREFASHYMIDYLRPGSQWIRYKNRTG
                      HVYMDGNFDTTTPVIRVLDPPIVASRIRFVPSSKNTRTVCMRAEIHGCKHEGVTYYST
                      VPDGSRLDTLDFKDSMFEDSQIYTESGIKRYGSWPANRRFRSAASPFEKNQMNNSWIG
                      WNRDTTDGRITILFEFEEVHNFTDVVLATFGNRIDGIDVIFSQDGKTFPLFSQISSSE
                      RQSLNNTSRRYDFRVPLHNRAGRKVRISIKFSSDWMFLTEVHFTSAANLTLLSEKIPP
                      PSSAATQQLLVVCGIIFLTIFACVAYCVSVCLKRRQKNKSVDSNVKKDLIITHMGNKP
                      TCHVFPSNGKLSNGHYEVANDILYARSQKSTLLSVSSKSTFSCRAIPPTWTDFNFPPP
                      PEGREEHTYSQPVSPENSSNGSYKSVRKIQALKKYPSSALLIGKAIGEGKFTMIKECI
                      IFGGLKCAHKSTKEEDCVHGTRALGDEIACLLQCGRHPRIVELFGVDESYNLLLEHVE
                      YGCIRNFWMASEAPLDTEFLVRICKDIYSAMAYLESIRIVHGHFTPNNILMDGEFHAK
                      VCSPRGPSHHAQLRYSAPESIVNNEFTHKSDAWAVATTVYEMAYQCRQRPYEELTNEQ
                      IVDNACALLDHQPNAVVPLMPTVFNYEILQLLTRCFRVDQLERPTFERLVKPFQD"
     gene             22629. .27352
                      /gene="C25F6.3"
     CDS              join(22629. .22778,22996. .23109,23156. .23244,23886. .24135,
                      24185. .24653,24703. .24875,24923. .25050,25114. .25196,
                      25344. .25495,25544. .26161,26225. .26956,27003. .27136,
                      27190. .27352)
                      /gene="C25F6.3"
                      /note="coded for by C. elegans cDNA yk28e9.3; coded for by
                      C. elegans cDNA yk28e9.5; coded for by C. elegans cDNA
                      yk28h9.3; coded for by C. elegans cDNA yk28h9.5; coded for
                      by C. elegans cDNA yk40e4.5; coded for by C. elegans cDNA
                      yk5h3.5;  weakly similar to dihydroorotate dehydrogenase
                      and glutamate synthase"
                      /codon_start=1
                      /protein_id="AAA80433.1"
                      /db_xref="GI:1049458"
                      /translation="MSYDFYAFCLSIFIFCKFFNIIFTAMTPKPNTTSPTNNLPLLSK
                      DSPDIESLLILNPKVQDKANAVPSAVTKKNKHNWKRNEEKGCGSTCGESKLKNDFRDI
                      KHTTLSERGALKEAMRCLKCADAPCQKSCPTQLDVKSFITSISNKNYYGAARQILSDN
                      PLGLTCGMICPTSDLCVGSCNLQASEEGAINIGGLQQYACDVFKQMNVRQIVSKEVRE
                      NRNASHKEQVALIGCGPASISCASFLARLGYTDITIYEKRAYIGGLSSAEIPQFRLPY
                      DVVDFEIQLARDIGVQIETNRPLGKDGLTLAKLKEQGAAAVFIGIGNPEPKIDPLFEG
                      LTIENGFYTSKNYLPAVAAASKPGMCGCKRTPLPTMRGRVVVLGAGDTAMDCATSALR
                      CGASRVTIAFRKGFTGIRAVPEEMEAAKEEKCEFLPFSAPRKINVKDGRIVSIEFNKT
                      EQDDNGKWYEDEEQIVILKCDYVISAFGSTLKEDAVLSALQPCQLNKWGGIEVDSTTQ
                      QTSEKWVFAGGDVAGVAETTVESVNDGKIAAWNMHRYIQSLHGNQVSETPELPQFFTP
                      IDEVDISVDMCGVKFENPFGLASAPPTTSGPMCRRAFEQGWGFILTKTYGLDKDLVTN
                      VSPRIVRGSTSGPLYGPNQGSFMNIELISEKSCEYWLQCIRELKRDHPTKIVIASIMC
                      VYNKADWIELATKSEEAGADILELNLSCPHGMGEKGMGLACGQSPEIVKEICRWVRAC
                      VKIPFFPKMTPNITDVREIARAARDGGASGVTATNTVSSLMHMKADGNAWPAIGSTKR
                      TTYGGMSGSAIRPIAMKAVSSIANELDGFPIMATGGIESAETGLGFLMAGASVLQVCS
                      AVQNQDFTVVDDYCTGLKALLYLSGAESLKNWDGQSPPIEKHQKGKPILLQGQKKMPF
                      FGKYRDEREKLEAIKLSESNLLDTENYHFASRPDTQVSRVPTVEDVIGKALPRIGPYV
                      TLDNQEQKVAIIDDDMCINCGKCYMTCNDSGYQAITFDPVTHQPHVTEDDCTGCTLCY
                      SVCPIPECIEMVPRTGPWKAPKRGVKPSVEPGTPKVVKVDQRGRVILDTTGGMQ"
     gene             34301. .39725
                      /gene="C25F6.2"
     CDS              join(34301. .34613,35496. .35812,35889. .36107,36155. .36251,
                      36299. .36389,36443. .36794,36851. .37157,37216. .37684,
                      37743. .38266,38350. .38632,38680. .38801,38857. .38870,
                      38931. .38949,39228. .39253,39483. .39725)
                      /gene="C25F6.2"
                      /note="coded for by C. elegans cDNA yk25e5.3; coded for by
                      C. elegans cDNA yk25e5.5;  similar to repeat guanylate
                      kinase domain of D. melanogaster lethal(1) discs large-1
                      tumor suppressor protein (SP:DLG1_DROME, P31007) and R.
                      norvegicus postsynaptic density protein 95 (PSD-95)
                      (SP:PSD9_RAT, P31016)"
                      /codon_start=1
                      /protein_id="AAA80434.1"
                      /db_xref="GI:1049459"
                      /translation="MAACLCVCVCVCACVCRKFLFLSDKLCLIAAQFFLFVFNASFLP
                      SSVFRLLEALKPVVYFYFQSQKLVDQFNFFNFQRPASKENRSEWEERLTEEGSMSHES
                      SEKAHKAIENVEDYCQTLTRHGNEELRTNLERVITTFKSNLMHSLLDIHDLYEQTLLS
                      ERKSDAEKNMEVRRVIERLEGGPHSYNSRPAATTSTSNYNLSSTTPLISDLRDRGGFS
                      YLNGGGLGNGLGNGLGNGLLSSPYNSSSTHYLHERQRQTSHDGTWRETTTRTVDTPSG
                      LERRVVEHTGVIDDHGRKWELENIVLEKGHTGLGFSITGGMDQPTEDGDTSIYVTNII
                      EGGAALADGRMRKNDIITAVNNTNCENVKHEVAVNALKSSGNVVSLSLKRRKDEAFLP
                      IGGNFGGSTSYLRSGVTPSVSAGNLQHAIHSPSAPIHPPPPPPVHHGSLSQLSVGQYR
                      STRPNTSVIDLVKGARGLGFSIAGGQGNEHVKGDTDIYVTKIIEEGAAELDGRLRVGD
                      KILEVDHHSLINTTHENAVNVLKNTGNRVRLLIQQGTGAIFNDSASQQFMPTTPILRP
                      SSVQDYNRSQMGSQSHLSYGGPLNTSYSSQAPIAIPLEPRPVQLVKGQNGLGFNIVGG
                      EDNEPIYISFVLPGGVADLSGNVKTGDVLLEVNGVVLRNATHKEAAEALRNAGNPVYL
                      TLQYRPQEYQIFESKIEKLRNDVIAQSRMGTLSRKSEYVRALFDYDPSRENSVAPHRS
                      MGFNYGDILHIINSSDDEWWTARKVHENGEETAEGVIPSKKRVEKRERLRRKQVNFNS
                      GSQSLGRNSSTTGLENRRGSRSQLSFSRKFPFVKSTDRLNDLNEESSNVAEEPVWSYQ
                      AVEQQAINYVRPVIILGALKDRINDELVNRDPSKFSSCVPHTSRPPREGEVNGRDYYF
                      VNKHNMEEDVKNNLFIEAGQFQNNLYGTSIQSVRDVANQGRHCILDVSGNAIRRLQSN
```

```
                     ANIQPISIFIKPSSAQQILELDSQLATNRQDDRAMSGEEAQAQYSRCHRIEQTFGDLF
                     TQEISNTPIWVPQTLLLELRNLSFLLPSPSSPNSIHFLLEIVTVCDILNASFVSTNRK
                     SLSYSRLSTDEKLRLISIFWDFYYSESILFSSIGSKLFFFNYF"
     tRNA            39962. .40034
                     /note="Met;  codon recognized: ATG"
     tRNA            40305. .40377
                     /note="Met;  codon recognized: ATG"
     tRNA            40596. .40668
                     /note="Met;  codon recognized: ATG"
     gene            41327. .42523
                     /gene="C25F6.1"
     CDS             join(41327. .41484,42079. .42208,42260. .42523)
                     /gene="C25F6.1"
                     /note="coded for by C. elegans cDNA CEESH77F"
                     /codon_start=1
                     /protein_id="AAA80435.1"
                     /db_xref="GI:1049460"
                     /translation="MAKMKMTQEELVQAFNQYMPKRPVYPPLVHDDDVLAHSATESEG
                     SDEEEDDLTKRLKVTLAFSTDMMSECRTAWEDSTCKPPIQYLEDNIQKDVEYILEKLV
                     NLKIKPNNPIEWLGYCLLNMNINDPRKSKDHKNKPSELLRDFKIEALVNMVWNMEEGG
                     SGSISTLHKQAPSPTAGKRRNPS"
BASE COUNT    13561 a   8218 c   7586 g  14438 t
ORIGIN


  Query Match             8.4%;  Score 72;  DB 6;  Length 43803;
  Best Local Similarity   63.1%;  Pred. No. 0.00012;
  Matches  111;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy     260 TTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAACCACCTCTGGC 319
           ||  |  ||||| | || || ||||| || ||    ||  |||| ||| | |||  |||
Db   25784 TTGGACAAGGACTTGGTCACCAATGTCTCTCCACGTATTGTCCGAGGATCTACCAGTGGA 25843

Qy     320 CCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCAGTGAGAAAACG 379
           ||  | |||||  ||    ||| | || ||  ||||||||||||| |||   ||||||
Db   25844 CCACTTTATGGACCAAACCAAGGATCATTCATGAATATTGAGCTTATCTCAGAGAAATCC 25903

Qy     380 GCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAGACAACaT 435
             || ||| |||  |||| || | |  ||  | ||   ||||   |||   || ||
Db   25904 TGTGAATACTGGCTTCAATGTATTAGGGAGTTGAAACGTGACCATCCAACTAAAAT 25959


RESULT  13
AC014157
LOCUS       AC014157    24429 bp    DNA             HTG       16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC014157
VERSION     AC014157.1  GI:6437178
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 24429)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10214237 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .24429
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     6835 a   5227 c   5359 g   7008 t
ORIGIN


  Query Match             8.3%;  Score 71.4;  DB 63;  Length 24429;
  Best Local Similarity   63.9%;  Pred. No. 0.00016;
  Matches  108;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy     267 AGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAACCACCTCTGGCCCCATGT 326
           ||||  | || || ||||| || ||    ||| || |||| ||||| || |||  || ||
Db   14044 AGGATCTGGTCACGAATGTCTCGCCGCGCATCGTCAGGGGCACCACGTCGGGCTACAAGT 14103

Qy     327 ATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCAGTGAGAAAACGGCTGCAT 386
           |||| ||    ||  |||  || ||||| || || || |||   |||||   ||| |  |
Db   14104 ATGGACCGCAGCAGGGCTGTTTCCTGAACATCGAACTAATCTCGGAGAAGCGGGCCGAGT 14163

Qy     387 ATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAGACAACaT 435
           | |||    ||    |    ||||| |||  ||||||||| || || ||
Db   14164 ACTGGCTGAAATCGATTGGAGAACTGAAGCGTGACTTCCCCGAGAAGAT 14212


RESULT  14
AE003446/c
LOCUS       AE003446   303092 bp    DNA             INV       04-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386054 section 30
            of 35, complete sequence.
ACCESSION   AE003446 AE002566
VERSION     AE003446.2  GI:10728536
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 303092)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
```

```
REFERENCE   2  (bases 1 to 303092)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7290992.
FEATURES             Location/Qualifiers
     source          1. .303092
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="X"
     mRNA            join(<6691. .6699,8631. .>8972)
                     /gene="CG15364"
                     /product="CT35383"
                     /db_xref="FLYBASE:FBan0015364"
                     /db_xref="FLYBASE:FBgn0030075"
                     /evidence=not_experimental
     gene            <6691. .>8972
                     /gene="CG15364"
                     /map="8C7-8C9"
                     /db_xref="FLYBASE:FBan0015364"
                     /db_xref="FLYBASE:FBgn0030075"
                     /evidence=not_experimental
     CDS             join(6691. .6699,8631. .8972)
                     /gene="CG15364"
                     /note="CG15364 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0015364"
                     /db_xref="FLYBASE:FBgn0030075"
                     /evidence=not_experimental
                     /protein_id="AAF46431.1"
                     /db_xref="GI:7290993"
                     /translation="MYMSLRSSARYAKAKSKSAINLNAHLKKQQKLEEGHPEKKRQDK
                     KILRQSVSEDQPIYANVDEVIPIHLEPETGTEVDALMGNAQVSGPKNPGTQLQTTSSI
                     CIRLHCIPSNKIPN"
     mRNA            join(<17819. .19495,19597. .19682,19745. .19970,20028. .20777,
                     20838. .20948,21010. .21879)
                     /gene="CG4040"
                     /product="CT13364"
                     /db_xref="FLYBASE:FBan0004040"
                     /db_xref="FLYBASE:FBgn0030076"
     gene            <17819. .>21879
                     /gene="CG4040"
                     /map="8C8-8C9"
                     /db_xref="FLYBASE:FBan0004040"
                     /db_xref="FLYBASE:FBgn0030076"
     CDS             join(17819. .19495,19597. .19682,19745. .19970,20028. .20777,
                     20838. .20948,21010. .21837)
                     /gene="CG4040"
                     /note="CG4040 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0004040"
                     /db_xref="FLYBASE:FBgn0030076"
                     /protein_id="AAF46432.1"
                     /db_xref="GI:7290994"
                     /translation="MNHLRVHKFKLGGKDNAAMADLPQSENTTNNNNNIGSNNNQKGG
                     AKGFLGRIKRYSVLGNKLGRRNLGSMNLNPPNNVVDNSVSPTKLHTGSYNMTGSHSDD
                     HRLEIGAPILISTTTLDTDRFDVTEARLKQIGGGIAQTSSIVRTLTPRSSDEEEFVDA
                     RSSPQQMERDEDLQIADDQVEKNDRDQYGHQFQLPKKQHREEEMGDLKYLDDLEEPEQ
                     FETPKHEVEEKLLTANQDVKRPITPITIAIDPPIRRPRIARQQRQSQSVQNLHRSELK
                     VYLHKSNSVALDMNVTAPAKLGLDLNLPELPELYGASELSLAASDNKENLPAVKPSAE
                     RSPIPNGGFMAQQPHFKSLDSFHLNTRSHSNQSSQQSSRQSLACSSSNGEYDFDLKSV
                     SYQSLNAQNLFVSIDELQELTRQINETEDFGKEIDLEYCAHRDQLRPSERRITLLKNK
                     NQTLINFNHNKEKLRKGWHGMKHWLGEEGSKIKEAVRQQTPLKRLAQSRSNLNQSTAD
                     ASRQSMSPERTQRDVTESCEDVTERTEMESSMSHRNSEEDLSPHAKRFKDEGQNGFEE
                     LRRYVKQGGDFSKELIFVLQERADSELIYSKSLSKLANKLNKAGREIPGSVADAWRGV
                     ATEMESRSDIHRQLAASLTDELVKPLKIVVEGHHKARKAVESNVDKAARVLGEWRASE
                     AKAKKASHTAARENEKLQDAMLDVRIQKSPSIALLHQGPNKQAAEKELKSAEKDCVKL
                     DNKRKKAEEAVKRADVEYYTLCVRSERARVDWEMAVLRGSAQLQSSEQQRLGNMHNFA
                     QQYARLISDMNPILGGLSTRLQPQLDACNVAKDMQVVRHIRRNSEGPSEQLLPDFYCE
                     HTTLAMNRERRKHALIKLLQLVKTDLERERRSRDGLRGLSQSLNHQEHQNITDKLYHI
                     RSMLTYLEGARLKLHSALLELDHKPRATHPLAQHIHITRDRTGLQQSILKVPNWLKNN
                     DKSQTQQSTSMLAHDITDITTNHEDELPDENCSHLHSSSHSNNNSNGSCTDVSSVVKP
                     FNRSKSNIETNFTSQPKIISTTNAAIAVSVKSKTLANLQDGHHNNQQHHHHHHNHHSD
                     RGQADGGSNQQDSDFDEFSSQDEDDEPQQPQQQPQPLQSQQQQQQPLLQNPTINGQVQ
                     TQHFYQNAQELQKGQKNGSVPILGRCKALYSYTPKLYDELELSPGDIIEVHAKQDDGW
                     WLGALRNHIGIFPATYVEECA"
     mRNA            complement(join(<23607. .24527,24584. .24901,25007. .25148,
                     25214. .25346,25568. .25910,26018. .>26626))
                     /gene="CG15365"
                     /product="CT35384"
                     /db_xref="FLYBASE:FBan0015365"
                     /db_xref="FLYBASE:FBgn0030077"
                     /evidence=not_experimental
     gene            complement(<23607. .>26626)
                     /gene="CG15365"
                     /map="8C9-8C10"
                     /db_xref="FLYBASE:FBan0015365"
                     /db_xref="FLYBASE:FBgn0030077"
                     /evidence=not_experimental
     CDS             complement(join(23607. .24527,24584. .24901,25007. .25148,
                     25214. .25346,25568. .25910,26018. .26626))
                     /gene="CG15365"
                     /note="CG15365 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0015365"
                     /db_xref="FLYBASE:FBgn0030077"
                     /evidence=not_experimental
                     /protein_id="AAF46433.1"
                     /db_xref="GI:7290995"
                     /translation="MSAATHMAMPAHAVVPPPGTPPVLGGSGCSNLSLPSGKRGHKRS
                     VSLENNAPLALRSTPSMGQSSLGSPLLQFGQAAPGQGHSQGFHTGTLKSRQEQRRLSQ
                     KFGSHSNLDDEGVGIGVGMGMGMGLQMRSKFQNIRQMFELSRSCVGGGGGDGRGGAAE
                     EEAMQSLPATLPQQHQQQQQRRTTPIAGGSRHQQQQQMGEAEQTSGNFLRPIAFKPIP
                     FEPDYRIACQQQQQHQQQLQLQQQQQQQQQQQQQQQQQLQLQHQQQQQHQQQHQHQHQQQ
                     HQSLQLPTEGGQAGGVAAERYGYGSTPSLAPLPTASAQKFGSTTDLRHLAARRRNHRL
                     LQRCSNEDALTLDLLSTGSHDRPDGASSKSSGTVVGSSDLTPSPSDSGISELEAALKD
                     RDSELSYLRQAMEHNEKDKEVYWEHETQRLRLFYEGQQRECQLKLRKMEQLLGLQQFQ
                     LKQHKLRQSEQVNRLQQQLDLARDSSQGLQQQVDALRHQLEDSEWRVCERNGEIALLK
                     TQLKEAHLEINMKDHAIVSLKHSSNRSSSSSNSNSNKSCDTNTSQSQPKQQAKEEQQQ
                     QQHPELQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQLKQQHQPDQQQLQKI
                     ITLKDQVIGALTSELAKLRKELSDLAIAHEYGEEPCGRYTRLKQQLDNLNEICQKTRK
                     YTTTSASPTAVASPAAQVAPPAPSPAAHVATPSCAQELPKLDVLVQRLQLDQGQGTQQ
                     TLDTLIEESTTYAEDIAKLRQKLDDFRLNLELEKRQWSAEKDKVLGYQKQLQAHYIQM
                     YQKLRCLDSAGAAVGQVEATSGN"
     mRNA            complement(38162. .38842)
                     /gene="CG10970"
                     /product="CT30739"
                     /db_xref="FLYBASE:FBan0010970"
                     /db_xref="FLYBASE:FBgn0030078"
     gene            complement(<38162. .>38842)
                     /gene="CG10970"
                     /map="8C9-8C12"
                     /db_xref="FLYBASE:FBan0010970"
                     /db_xref="FLYBASE:FBgn0030078"
     CDS             complement(38314. .38625)
                     /gene="CG10970"
                     /note="CG10970 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0010970"
                     /db_xref="FLYBASE:FBgn0030078"
                     /protein_id="AAF46434.1"
                     /db_xref="GI:7290996"
                     /translation="MAGQKEEILCCDFEVRGRVPKDAFELFALVQANALGLRGFIAQV
                     SEEKFKGRLEGEGKVIDAFKKLILAAADYVQAIKEFIIKNLKVIQEYTYKTFEIKVNL
                     K"
     mRNA            complement(join(<40014. .40361,40432. .40549,41242. .41250))
                     /gene="CG7267"
                     /product="CT22403"
                     /db_xref="FLYBASE:FBan0007267"
                     /db_xref="FLYBASE:FBgn0030079"
     gene            complement(<40014. .>41250)
                     /gene="CG7267"
                     /map="8C9-8C12"
                     /db_xref="FLYBASE:FBan0007267"
                     /db_xref="FLYBASE:FBgn0030079"
```

```
     CDS             complement(join(40144. .40361,40432. .40549,41242. .41250))

  Query Match             8.3%;  Score 71.4;  DB 4;  Length 303092;
  Best Local Similarity   63.9%;  Pred. No. 0.00016;
  Matches  108;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy     267 AGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAACCACCTCTGGCCCCATGT 326
           ||||  | || || ||||| || ||    ||| || |||| ||||| || ||| || ||
Db  165897 AGGATCTGGTCACGAATGTCTCGCCGCGCATCGTCAGGGGCACCACGTCGGGCTACAAGT 165838

Qy     327 ATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCAGTGAGAAAACGGCTGCAT 386
           |||| ||    ||  ||| || ||||| || || || |||   |||||   ||| |  |
Db  165837 ATGGACCGCAGCAGGGCTGTTTCCTGAACATCGAACTAATCTCGGAGAAGCGGGCCGAGT 165778

Qy     387 ATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAGACAACaT 435
           | |||    ||    |    ||||| |||  ||||||||| || || ||
Db  165777 ACTGGCTGAAATCGATTGGAGAACTGAAGCGTGACTTCCCCGAGAAGAT 165729


RESULT  15
DMU65491
LOCUS       DMU65491     2348 bp    mRNA            INV       26-SEP-1996
DEFINITION  Drosophila melanogaster Dreg-3 protein mRNA, complete cds.
ACCESSION   U65491
VERSION     U65491.1  GI:1561729
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 2348)
  AUTHORS   Van Gelder,R.N., Bae,H., Palazzolo,M.J. and Krasnow,M.A.
  TITLE     Extent and character of circadian gene expression in Drosophila
            melanogaster: identification of twenty oscillating mRNAs in the fly
            head
  JOURNAL   Curr. Biol. 5 (12), 1424-1436 (1995)
  MEDLINE   96362140
REFERENCE   2  (bases 1 to 2348)
  AUTHORS   Van Gelder,R.N., Bae,H., Palazzolo,M.J. and Krasnow,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1996) Ophthalmology and Visual Sciences,
            Washington University School of Medicine, 660 S. Euclid Avenue, Box
            8096, St. Louis, MO 63110, USA
FEATURES             Location/Qualifiers
     source          1. .2348
                     /organism="Drosophila melanogaster"
                     /strain="Canton-S"
                     /db_xref="taxon:7227"
     CDS             358. .2238
                     /note="similar to dihydropyrimidine dehydrogenase;
                     rhythmically expressed gene 3"
                     /codon_start=1
                     /product="Dreg-3 protein"
                     /protein_id="AAC47288.1"
                     /db_xref="GI:1561730"
                     /translation="MEFCRTEQNENDEWVEDEEQTQRLKANFVISAFGSGLEDQDVKA
                     ALAPLQFRGELPVVDRVTMQSSVKQVFLGGDLAGVANTTVESVNDGKVAAWSIHCQLQ
                     GLPLDTPAALPLFYTDIDAVDISVEMCGIRFENPFGLASAPPTTSTAMIRRAFEQGWG
                     FVVTKTFGLDKDLVTNVSPRIVRGTTSGYKYGPQQGCFLNIELISEKRAEYWLKSIGE
                     LKRDFPEKIVIASIMCSFNEEDWTELAIKAEQSGADALELNLSCPHGMGERGMGLACG
                     QDPELVEQISRWVRKAVKLPFFIKLTPNITDIVSIAAAAKREEPMRSAINTVQGLMGL
                     KADSTAWPAIGKEQRTTYGGVSGNATRPMALKAISDIANRVPGFPILGIGGIDSGEVA
                     LQFIHAGATVLQICSSVQNQDFTVIEDYCTALKALLYLKRIRHQSMVPSGMASHHPRR
                     SIRRQARCPFDRRGKATLGFFGPYQRQRDIKMAELRSQKGALSWDAEQVKATPPASNG
                     APNPAPRIKDVIGAALDKIGSYNKLDNKQQKVALIDDDMCINCGKCYMTCADSGYQAI
                     EFDKDTHIPHVNDDCTGCTLCVSVCPIIDCITMVPKKIPHVIKRGVEEKIFYTHALSQ
                     CQ"
BASE COUNT      538 a    649 c    684 g    477 t
ORIGIN

  Query Match             8.1%;  Score 69.8;  DB 6;  Length 2348;
  Best Local Similarity   63.3%;  Pred. No. 0.00034;
  Matches  107;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy     267 AGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAACCACCTCTGGCCCCATGT 326
           ||||  | || || ||||| || ||    ||| || |||| ||||| || ||| || ||
Db     869 AGGATCTGGTCACGAATGTCTCGCCGCGCATCGTCAGGGGCACCACGTCGGGCTACAAGT 928

Qy     327 ATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCAGTGAGAAAACGGCTGCAT 386
           |||| ||    ||  | |  || ||||| || || || |||   |||||   ||| |  |
Db     929 ATGGACCGCAGCAGGGTTGTTTCCTGAACATCGAACTAATCTCGGAGAAGCGGGCCGAGT 988

Qy     387 ATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAGACAACaT 435
           | |||    ||    |    ||||| |||  ||||||||| || || ||
Db     989 ACTGGCTGAAATCGATTGGAGAACTGAAGCGTGACTTCCCCGAGAAGAT 1037


Search completed: November 20, 2001, 16:51:17
Job time: 6131 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2001, 15:32:06 ; Search time 124.09 Seconds
(without alignments)
4356.703 Million cell updates/sec

Title: US-09-308-080-1
Perfect score: 861
Sequence: 1 TGTTAATGAAGATAAATATT.....AGTGGGAAATAATTTATTAA 861

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601:*
1: /SIDS8/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	856.2	99.4	861	18	AAT91798	Dihydropyrimidine
	2	172.8	20.1	3951	20	AAX22903	Human DPD DNA. Ho
	3	171.2	19.9	3957	17	AAT14077	Human dihydropyrim
	4	171.2	19.9	4368	16	AAT03133	Human dihydropyrim
	5	171.2	19.9	4369	16	AAT03143	Human dihydropyrim
	6	154.8	18.0	4414	16	AAT03132	Bovine dihydropyri
	7	154.8	18.0	4430	16	AAT03142	Bovine dihydropyri
	8	152.6	17.7	936	22	AAF58252	Oligonucleotide D1
	9	152.6	17.7	936	22	AAF58254	Oligonucleotide D1
	10	152.6	17.7	936	22	AAF58257	Oligonucleotide D1
	11	152.6	17.7	936	22	AAF58259	Oligonucleotide D2
	12	152.6	17.7	936	22	AAF58262	Oligonucleotide D2
	13	152.6	17.7	938	22	AAF58255	Oligonucleotide D1
	14	151.6	17.6	4447	17	AAT14078	Pig dihydropyrimid
c	15	148.6	17.3	936	22	AAF58252	Oligonucleotide D1
c	16	148.6	17.3	936	22	AAF58254	Oligonucleotide D1
c	17	148.6	17.3	936	22	AAF58257	Oligonucleotide D1
c	18	148.6	17.3	936	22	AAF58259	Oligonucleotide D2
c	19	148.6	17.3	936	22	AAF58262	Oligonucleotide D2
c	20	148.6	17.3	938	22	AAF58255	Oligonucleotide D1
	21	61.2	7.1	244	22	AAF58238	Oligonucleotide D1
c	22	59.8	6.9	244	22	AAF58238	Oligonucleotide D1
c	23	45	5.2	5852	12	AAQ11710	Dictyostelium plas
	24	42.4	4.9	3648	22	AAD02172	Plasmid pHiNk181 c
	25	42.4	4.9	3648	22	AAD02173	Plasmid pHiNk184 c
	26	42.4	4.9	11461	22	AAD02175	Plasmid pZU634 for
	27	42.4	4.9	12766	22	AAD02174	Plasmid pZU623 for
	28	42.4	4.9	13737	21	AAA54212	Transformation vec
c	29	42.2	4.9	2400	18	AAT49364	Nicotiana plumbagi
c	30	42.2	4.9	8318	20	AAX20264	Borrelia burgdorfe
c	31	41.4	4.8	745	20	AAX39751	Gastric cancer ass
c	32	41	4.8	3947	19	AAV69429	Ehrlichia sp. exte
c	33	41	4.8	3947	21	AAZ89994	Ehrlichia antigen
	34	40.2	4.7	2313	21	AAC76449	Human ORFX ORF2004
	35	40	4.6	37808	20	AAX02780	Vector pMVX-BG DNA
	36	39.6	4.6	81369	21	AAA97997	Human T gene DNA.
	37	39.4	4.6	639	21	AAZ98055	Human secreted pro
c	38	39.4	4.6	1717	20	AAX20311	Borrelia burgdorfe
c	39	39	4.5	14752	20	AAX20256	Borrelia burgdorfe
c	40	39	4.5	116277	20	AAX20249	Borrelia burgdorfe
	41	39	4.5	910715	20	AAX20248	Borrelia burgdorfe
c	42	38.8	4.5	5810	18	AAT91323	Arabidopsis thalia
	43	38.8	4.5	8911	20	AAX13588	Enterococcus faeca
	44	38.6	4.5	1400	20	AAZ27735	Human DNA marker c
	45	38.6	4.5	1591	21	AAZ95204	Human UGT2B7 exon

ALIGNMENTS

RESULT 1
AAT91798
ID AAT91798 standard; DNA; 861 BP.
XX
AC AAT91798;
XX
DT 14-APR-1998 (first entry)
XX
DE Dihydropyrimidine dehydrogenase genomic fragment.
XX
KW Human; dihydropyrimidine dehydrogenase; DPD; slicing defect;
KW detection; 5-fluorouracil; cancer; anticancer; uraciluria; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT exon 269..433
FT /*tag= a
FT /note= "Encodes amino acids 581-635 of the DPD protein"
XX
PN WO9735034-A1.
XX
PD 25-SEP-1997.
XX
PF 19-MAR-1997; 97WO-US04269.
XX
PR 20-MAR-1996; 96US-0013835.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Fernandez-Salguero P, Gonzalez FJ;
XX
DR WPI; 1997-480236/44.
DR P-PSDB; AAW30092.

```
XX
PT   Detecting a splicing defect in the di:hydro:pyrimidine dehydrogenase
PT   gene - used to identify subjects sensitive to 5-fluorouracil, toxic
PT   to individuals with DPD defects
XX
PS   Disclosure; Fig 1; 38pp; English.
XX
CC   A novel method has been developed for detecting the splicing defect in
CC   the dihydropyrimidine dehydrogenase (DPD) gene. The method comprises
CC   determining whether genomic DNA containing the DPD gene has a wild-type
CC   intron-exon boundary for an exon that encodes amino acids (aa) 581-635
CC   of the corresponding DPD protein. The present sequence represents a
CC   DPD genomic fragment which encodes the amino acids 581-635. The method
CC   is used specifically to determine sensitivity of subjects to the
CC   anticancer agent 5-fluorouracil, which is dangerously toxic to those
CC   with DPD defects. It can also be used to diagnose DPD-deficiency
CC   disorders such as uraciluria.
XX
SQ   Sequence 861 BP; 300 A; 140 C; 161 G; 260 T; 0 other;


  Query Match             99.4%;  Score 856.2;  DB 18;  Length 861;
  Best Local Similarity   99.7%;  Pred. No. 1.2e-181;
  Matches  858;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 TGTTAATGAAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 tgttaatgaagataaatatttttgttttttcgctgttctaaacctagggttacaagaagt 60

Qy      61 AATTTATCTGGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 aatttatctggagctaacaaatactttatttttacctttttatttgcaagtagtttatgtt 120

Qy     121 CAATTCTAATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 caattctaatttaatgtatattaaaaattcctctgcaaatatgtgaggagggacctcata 180

Qy     181 AAATATTGTCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 aaatattgtcatatggaaatgagcagataataaagattatagcttttctttgtcaaaagg 240

Qy     241 AGACTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCAT 300
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     241 agactcaatatctttactctttcatgaggacattgtgacaaatgtttcccccagaatcat 300

Qy     301 CCGGGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ccggggaaccacctctggccccatgtatggccctggacaaagctcctttctgaatattga 360

Qy     361 GCTCATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 gctcatcagtgagaaaacggctgcatattggtgtcaaagtgtcactgaactaaaggctga 420

Qy     421 CTTCCCAGACAACaTAAGTGTGATAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTG 480
           |||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db     421 cttcccagacaacggaagtgtgataaaaatctaaaacaagagaattggcataagttggtg 480

Qy     481 AATGTTTATTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCAA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 aatgtttatttaaacatccaattcataggcttataaatattaatgtgtatattttatcaa 540

Qy     541 CGAATCTGCCAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 cgaatctgccagttgctttgctgatgcatagaaagataaaaaagaaagaaaagctcaaga 600

Qy     601 ACTCATAAAAACCCACACAATGTGAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAA 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 actcataaaaacccacacaatgtgaagctctgttataaatgggtgccatgtaagatggaa 660

Qy     661 GAAGTATCTACATAAGCAGAAGGAAGAGAAATGAAATACTCATTTTATTGAGTTGGCCCC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 gaagtatctacataagcagaaggaagagaaatgaaatactcattttattgagttggcccc 720

Qy     721 CACTGTATGTGGCTGGTATTTATGAAGGTGATGACCCAGGAAGAAATTGTAAACTATAAA 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 cactgtatgtggctggtatttatgaaggtgatgacccaggaagaaattgtaaactataaa 780

Qy     781 CCACTCCAAATATAAACCCGAGGCAGAAGCAGCATATCTCCTATGAAGCCTGTATTTACT 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 ccactccaaatataaacccgaggcagaagcagcatatctcctatgaagcctgtatttact 840

Qy     841 CAGTGGGAAATAATTTATTAA 861
           |||||||||||||||||||||
Db     841 cagtgggaaataatttattaa 861


RESULT   2
AAX22903
ID   AAX22903 standard; DNA; 3951 BP.
XX
AC   AAX22903;
XX
DT   28-MAY-1999  (first entry)
XX
DE   Human DPD DNA.
XX
KW   DPD; dihydropyrimidine-dehydrogenase; monoclónal antibody; MAb;
KW   immunoassay reagent; cancer patient; treatment; antitumour agent;
KW   5-fluorouracil; affinity purification; toxicity; ss.
XX
OS   Homo sapiens.
XX
PN   DE19837391-A1.
XX
PD   25-FEB-1999.
XX
PF   18-AUG-1998;   98DE-1037391.
XX
PR   22-AUG-1997;   97EP-0114630.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI   Hasegawa M,  Yoshikubo T;
XX
DR   WPI; 1999-155202/14.
DR   P-PSDB; AAW93361.
XX
PT   Monoclonal antibody specific for dihydropyrimidine dehydrogenase -
PT   for assessing patient response to 5-fluorouracil antitumor agents
XX
PS   Disclosure; Page 15-17; 34pp; German.
XX
CC   This invention describes a monoclonal antibody (MAb) specific for
CC   dehydropyrimidine dehydrogenase (DPD). This MAb is used as immunoassay
CC   reagents to identify a lack of DPD in a patient and to assess the
CC   sensitivity of cancer patients to treatment with antitumor agents of the
CC   5-fluorouracil (5-FU) type. The MAb can also be used for affinity
CC   purification of DPD. DPD is involved in reduction of 5-FU (and related
CC   catabolites and derivatives) and lack of it is associated with increased
CC   toxicity of this type of antitumor agent. It has specific binding
CC   interaction. The MAb provide a sensitive and reliable test for DPD,
CC   which is simple, rapid and suitable for routine screening.
XX
SQ   Sequence 3951 BP; 1153 A; 785 C; 896 G; 1117 T; 0 other;


  Query Match             20.1%;  Score 172.8;  DB 20;  Length 3951;
  Best Local Similarity   93.8%;  Pred. No. 1.5e-29;
  Matches  180;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy     244 CTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCG 303
           | | | | ||||     || || |||||||||||||||||||||||||| |||||||||
Db    1797 caccaaaactttctctcttgataaggacattgtgacaaatgtttcccccagaatcatccg 1856

Qy     304 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1857 gggaaccacctctggccccatgtatggccctggacaaagctcctttctgaatattgagct 1916

Qy    364 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 423
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1917 catcagtgagaaaacggctgcatattggtgtcaaagtgtcactgaactaaaggctgactt 1976

Qy    424 CCCAGACAACaT 435
          ||||||||||||
Db   1977 cccagacaacat 1988


RESULT   3
AAT14077
ID   AAT14077 standard; cDNA; 3957 BP.
XX
AC   AAT14077;
XX
DT   11-JUN-1996  (first entry)
XX
DE   Human dihydropyrimidine dehydrogenase cDNA.
XX
KW   Dihydropyrimidine dehydrogenase; DPD; DPYD gene; probe; deficiency;
KW   5-fluorouracil; cytostatic; cancer; selectable marker; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             88..3165
FT                   /*tag=  a
FT   primer_bind     complement (36..55)
FT                   /*tag=  b
FT                   /note= "primer RTF1"
FT   primer_bind     1536..1558
FT                   /*tag=  c
FT                   /note= "primer RTR1"
FT   primer_bind     complement (1539..1558)
FT                   /*tag=  d
FT                   /note= "primer H13"
FT   primer_bind     2426..2445
FT                   /*tag=  e
FT                   /note= "primer RTR4"
FT   primer_bind     complement (2424..2447)
FT                   /*tag=  f
FT                   /note= "primer RTR5"
FT   primer_bind     3320..3343
FT                   /*tag=  g
FT                   /note= "primer RTR5"
XX
PN   WO9608568-A2.
XX
PD   21-MAR-1996.
XX
PF   07-SEP-1995;   95WO-US12016.
XX
PR   12-SEP-1994;   94US-0304309.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Fernandez-Salguero P,  Gonzalez FJ;
XX
DR   WPI; 1996-179943/18.
DR   P-PSDB; AAR91420.
XX
PT   Pig and human di:hydro:pyrimidine dehydrogenase (DPD) genes and
PT   probes  - useful for detection of DPD deficiencies and
PT   identification of humans at risk of toxic reaction to
PT   5-fluoro:uracil anti-cancer treatment
XX
PS   Claim 1; Page 43-48; 78pp; English.
XX
CC   Human dihydropyrmidine dehydrogenase (DPD) cDNA (AAT14077) was
CC   isolated as 3 overlapping fragments obtd. from human liver
CC   cDNA libraries screened using pig DPD cDNA (see also AAT14077).
CC   The DPD gene, DPYD, is located at chromosome 1p22.  The cDNA
CC   can be inserted into a vector and used for prodn. of DPD (AAR91420)
CC   in transformed host cells, or used as a selectable marker that
CC   functions in prokaryotic and eukaryotic cells.  It can also be
CC   utilised as a probe to detect DPD deficiency in an individual,
CC   i.e. to identify risk of toxic reaction to 4-fluorouracil.
XX
SQ   Sequence 3957 BP; 1156 A; 787 C; 897 G; 1117 T; 0 other;


  Query Match             19.9%;  Score 171.2;  DB 17;  Length 3957;
  Best Local Similarity   93.2%;  Pred. No. 3.4e-29;
  Matches  179;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy    244 CTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCG 303
          | | | | ||||     || || ||||||||||||||||||||||||||| |||||||||
Db   1803 caccaaaactttctctcttgataaggacattgtgacaaatgtttcccccagaatcatccg 1862

Qy    304 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 363
          ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   1863 gggaaccacctctggccccatgtatgcccctggacaaagctcctttctgaatattgagct 1922

Qy    364 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 423
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1923 catcagtgagaaaacggctgcatattggtgtcaaagtgtcactgaactaaaggctgactt 1982

Qy    424 CCCAGACAACaT 435
          ||||||||||||
Db   1983 cccagacaacat 1994


RESULT   4
AAT03133
ID   AAT03133 standard; DNA; 4368 BP.
XX
AC   AAT03133;
XX
DT   04-JUN-1996  (first entry)
XX
DE   Human dihydropyrimidine dehydrogenase gene.
XX
KW   Bovine; liver; human; dihydropyrimidine dehydrogenase; DPD;
KW   pyrimidine catabolism; 5,6-dihydropyrimidine; pyrimidine analogue;
KW   fluoropyrimidine; anticancer drug; 5-fluorouracil; FUra; cancer;
KW   frameshift mutation; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             49..3126
FT                   /*tag=  a
FT                   /product= DPD
XX
PN   WO9528489-A1.
XX
PD   26-OCT-1995.
XX
PF   13-APR-1995;   95WO-US04567.
XX
PR   13-APR-1994;   94US-0227357.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Cheng X,  Diasio RB,  Johnson M,  Lu Z,  Zhang R;
XX
DR   WPI; 1995-373803/48.
DR   P-PSDB; AAR86001.
XX
PT   Novel di:hydro:pyrimidine dehydrogenase gene - used to optimise
PT   5-fluoro:uracil doses given to cancer patients
XX
PS   Claim 10; Page 140-56; 207pp; English.
XX
```

```
CC   This sequence represents the DNA sequence encoding human
CC   dihydropyrimidine dehydrogenase (DPD).  DPD catalyses the initial
CC   and rate limiting step in pyrimidine catabolism, the reduction of
CC   pyrimidines to 5,6-dihydropyrimidines.  DPD is a complex enzyme
CC   consisting of two identical subunits, containing FMN, FAD and iron-
CC   sulphur centers, and utilising NADPH as a cofactor.  DPD has also been
CC   shown to catalyse the reduction of various pyrimidine analogues
CC   including the fluoropyrimidine anticancer drug 5-fluorouracil (FUra).
CC   Up to 85% of administered FUra may be catabolised by DPD, and it
CC   therefore governs the effectiveness of FUra as an anticancer drug.
CC   DPD genes or fragments of them may be used in the detection of DPD in a
CC   sample, esp. isolated from a cancer patient.  According to the amount
CC   of DPD detected, a therapeutically effective amount of FUra may be
CC   determined and administered.  DPD deficiency, leading to life-
CC   threatening toxicity on exposure to FUra, in a human caused by a
CC   frameshift mutation may be determined by means of a molecular biological
CC   assay to detect the deletion of an A residue at codon 318 within the
CC   DPD-coding region.
XX
SQ   Sequence 4368 BP; 1325 A; 829 C; 932 G; 1282 T; 0 other;


  Query Match             19.9%;  Score 171.2;  DB 16;  Length 4368;
  Best Local Similarity   93.2%;  Pred. No. 3.5e-29;
  Matches  179;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy     244 CTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCG 303
           | | | | ||||     || || |||||||||||||||||||||||||| |||||||||
Db    1764 caccaaaactttctctcttgataaggacattgtgacaaatgtttcccccagaatcatccg 1823

Qy     304 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1824 gggaaccacctctggccccatgtatggccctggacaaagctcctttctgaatattgagct 1883

Qy     364 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 423
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1884 catcagtgagaaaacggctgcatattggtgtcaaagtgtcactgaactaaaggctgactt 1943

Qy     424 CCCAGACAACaT 435
           ||||||||||||
Db    1944 tccagacaacat 1955


RESULT   5
AAT03143
ID   AAT03143 standard; cDNA; 4369 BP.
XX
AC   AAT03143;
XX
DT   05-JUN-1996  (first entry)
XX
DE   Human dihydropyrimidine dehydrogenase cDNA.
XX
KW   Bovine; liver; human; dihydropyrimidine dehydrogenase; DPD;
KW   pyrimidine catabolism; 5,6-dihydropyrimidine; pyrimidine analogue;
KW   fluoropyrimidine; anticancer drug; 5-fluorouracil; FUra; cancer;
KW   frameshift mutation; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             49..3125
FT                   /*tag=  a
FT                   /product= Human lymphocyte DPD
FT   primer_bind     complement (36..60)
FT                   /*tag=  b
FT   primer_bind     complement (716..743)
FT                   /*tag=  c
FT   primer_bind     1260..1288
FT                   /*tag=  d
FT   primer_bind     3117..3139
FT                   /*tag=  e
FT   mutation        1000
FT                   /*tag=  f
FT                   /note= "Position of single base deletion in DPD
FT                          deficient patients"
XX
PN   WO9528489-A1.
XX
PD   26-OCT-1995.
XX
PF   13-APR-1995;   95WO-US04567.
XX
PR   13-APR-1994;   94US-0227357.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Cheng X,  Diasio RB,  Johnson M,  Lu Z,  Zhang R;
XX
DR   WPI; 1995-373803/48.
DR   P-PSDB; AAR86001.
XX
PT   Novel di:hydro:pyrimidine dehydrogenase gene - used to optimise
PT   5-fluoro:uracil doses given to cancer patients
XX
PS   Disclosure; Fig 7; 207pp; English.
XX
CC   This sequence represents the cDNA sequence encoding human dihydro-
CC   pyrimidine dehydrogenase (DPD).  This sequence was amplified using
CC   the primer sequences given in AAT03144-47.  DPD catalyses the initial
CC   and rate limiting step in pyrimidine catabolism, the reduction of
CC   pyrimidines to 5,6-dihydropyrimidines.  DPD is a complex enzyme
CC   consisting of two identical subunits, containing FMN, FAD and iron-
CC   sulphur centers, and utilising NADPH as a cofactor.  DPD has also been
CC   shown to catalyse the reduction of various pyrimidine analogues
CC   including the fluoropyrimidine anticancer drug 5-fluorouracil (FUra).
CC   Up to 85% of administered FUra may be catabolised by DPD, and it
CC   therefore governs the effectiveness of FUra as an anticancer drug.
CC   DPD genes or fragments of them may be used in the detection of DPD in a
CC   sample, esp. isolated from a cancer patient.  According to the amount
CC   of DPD detected, a therapeutically effective amount of FUra may be
CC   determined and administered.  DPD deficiency, leading to life-
CC   threatening toxicity on exposure to FUra, in a human caused by a
CC   frameshift mutation may be determined by means of a molecular biological
CC   assay to detect the deletion of an A residue at codon 318 within the
CC   DPD-coding region.
XX
SQ   Sequence 4369 BP; 1325 A; 829 C; 932 G; 1283 T; 0 other;


  Query Match             19.9%;  Score 171.2;  DB 16;  Length 4369;
  Best Local Similarity   93.2%;  Pred. No. 3.5e-29;
  Matches  179;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy     244 CTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCG 303
           | | | | ||||     || || |||||||||||||||||||||||||| |||||||||
Db    1764 caccaaaactttctctcttgataaggacattgtgacaaatgtttcccccagaatcatccg 1823

Qy     304 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1824 gggaaccacctctggccccatgtatggccctggacaaagctcctttctgaatattgagct 1883

Qy     364 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 423
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1884 catcagtgagaaaacggctgcatattggtgtcaaagtgtcactgaactaaaggctgactt 1943

Qy     424 CCCAGACAACaT 435
           ||||||||||||
Db    1944 tccagacaacat 1955


RESULT   6
AAT03132
ID   AAT03132 standard; DNA; 4414 BP.
XX
AC   AAT03132;
```

```
XX
DT   04-JUN-1996  (first entry)
XX
DE   Bovine dihydropyrimidine dehydrogenase gene.
XX
KW   Bovine; liver; human; dihydropyrimidine dehydrogenase; DPD;
KW   pyrimidine catabolism; 5,6-dihydropyrimidine; pyrimidine analogue;
KW   fluoropyrimidine; anticancer drug; 5-fluorouracil; FUra; cancer;
KW   frameshift mutation; ss.
XX
OS   Bos taurus.
XX
FH   Key             Location/Qualifiers
FT   CDS             68..3145
FT                   /*tag= a
FT                   /product= DPD
XX
PN   WO9528489-A1.
XX
PD   26-OCT-1995.
XX
PF   13-APR-1995;   95WO-US04567.
XX
PR   13-APR-1994;   94US-0227357.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Cheng X,  Diasio RB,  Johnson M,  Lu Z,  Zhang R;
XX
DR   WPI; 1995-373803/48.
DR   P-PSDB; AAR86000.
XX
PT   Novel di:hydro:pyrimidine dehydrogenase gene - used to optimise
PT   5-fluoro:uracil doses given to cancer patients
XX
PS   Claim 4; Page 115-32; 207pp; English.
XX
CC   This sequence represents the DNA sequence encoding bovine liver
CC   dihydropyrimidine dehydrogenase (DPD).  DPD catalyses the initial
CC   and rate limiting step in pyrimidine catabolism, the reduction of
CC   pyrimidines to 5,6-dihydropyrimidines.  DPD is a complex enzyme
CC   consisting of two identical subunits, containing FMN, FAD and iron-
CC   sulphur centers, and utilising NADPH as a cofactor.  DPD has also been
CC   shown to catalyse the reduction of various pyrimidine analogues
CC   including the fluoropyrimidine anticancer drug 5-fluorouracil (FUra).
CC   Up to 85% of administered FUra may be catabolised by DPD, and it
CC   therefore governs the effectiveness of FUra as an anticancer drug.
CC   DPD genes or fragments of them may be used in the detection of DPD in a
CC   sample, esp. isolated from a cancer patient.  According to the amount
CC   of DPD detected, a therapeutically effective amount of FUra may be
CC   determined and administered.  DPD deficiency, leading to life-
CC   threatening toxicity on exposure to FUra, in a human caused by a
CC   frameshift mutation may be determined by means of a molecular biological
CC   assay to detect the deletion of an A residue at codon 318 within the
CC   DPD-coding region.
XX
SQ   Sequence 4414 BP; 1315 A; 872 C; 956 G; 1271 T; 0 other;

  Query Match             18.0%;  Score 154.8;  DB 16;  Length 4414;
  Best Local Similarity   88.4%;  Pred. No. 1.5e-25;
  Matches  168;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy     246 CAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGG 305
           | | | ||||     || || |||| || ||||||||||||||| |||| ||||||||||
Db    1785 ccaaaactttctctcttgataaggatatagtgacaaatgtttcacccagaatcatccggg 1844

Qy     306 GAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCA 365
           | ||||||||||||||||||||||||||||||||||||||| || ||||||||||||||||
Db    1845 ggaccacctctggccccatgtatggccctggacaaagctctttcctgaatattgagctca 1904

Qy     366 TCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCC 425
           ||||||| ||||||||||||||||||||||||||||||||||||||| || ||||| |

Db    1905 tcagtgaaaaaacggctgcatattggtgtcaaagtgtcactgaactaaaagccgactttc 1964

Qy     426 CAGACAACaT 435
           ||||||| ||
Db    1965 cagacaatat 1974


RESULT   7
AAT03142
ID   AAT03142 standard; DNA; 4430 BP.
XX
AC   AAT03142;
XX
DT   04-JUN-1996  (first entry)
XX
DE   Bovine dihydropyrimidine dehydrogenase cDNA.
XX
KW   Bovine; liver; human; dihydropyrimidine dehydrogenase; DPD;
KW   pyrimidine catabolism; 5,6-dihydropyrimidine; pyrimidine analogue;
KW   fluoropyrimidine; anticancer drug; 5-fluorouracil; FUra; cancer;
KW   frameshift mutation; ss.
XX
OS   Bos taurus.
XX
FH   Key             Location/Qualifiers
FT   CDS             67..3144
FT                   /*tag= a
FT                   /product= Bovine DPD
FT   primer_bind     complement (2032..2051)
FT                   /*tag= b
FT                   /note= "Primer A"
FT   primer_bind     2080..2096
FT                   /*tag= c
FT                   /note= "Primer B"
FT   primer_bind     complement (2035..2078)
FT                   /*tag= d
FT                   /note= "Primer C"
FT   primer_bind     complement (85..104)
FT                   /*tag= e
FT                   /note= "Primer D"
FT   primer_bind     2141..2160
FT                   /*tag= f
FT                   /note= "Primer E"
FT   primer_bind     210..229
FT                   /*tag= g
FT                   /note= "Primer F"
XX
PN   WO9528489-A1.
XX
PD   26-OCT-1995.
XX
PF   13-APR-1995;   95WO-US04567.
XX
PR   13-APR-1994;   94US-0227357.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Cheng X,  Diasio RB,  Johnson M,  Lu Z,  Zhang R;
XX
DR   WPI; 1995-373803/48.
DR   P-PSDB; AAR86000.
XX
PT   Novel di:hydro:pyrimidine dehydrogenase gene - used to optimise
PT   5-fluoro:uracil doses given to cancer patients
XX
PS   Example 2; Figure 6; 207pp; English.
XX
CC   This sequence represents the cDNA sequence encoding bovine liver
CC   dihydropyrimidine dehydrogenase (DPD).  This sequence was isolated using
CC   the primers given in AAT03134-41 and AAT14638.  DPD catalyses the
CC   initial and rate limiting step in pyrimidine catabolism, the reduction
CC   of pyrimidines to 5,6-dihydropyrimidines.  DPD is a complex enzyme
CC   consisting of two identical subunits, containing FMN, FAD and iron-
```

CC sulphur centers, and utilising NADPH as a cofactor. DPD has also been
CC shown to catalyse the reduction of various pyrimidine analogues
CC including the fluoropyrimidine anticancer drug 5-fluorouracil (FUra).
CC Up to 85% of administered FUra may be catabolised by DPD, and it
CC therefore governs the effectiveness of FUra as an anticancer drug.
CC DPD genes or fragments of them may be used in the detection of DPD in a
CC sample, esp. isolated from a cancer patient. According to the amount
CC of DPD detected, a therapeutically effective amount of FUra may be
CC determined and administered. DPD deficiency, leading to life-
CC threatening toxicity on exposure to FUra, in a human caused by a
CC frameshift mutation may be determined by means of a molecular biological
CC assay to detect the deletion of an A residue at codon 318 within the
CC DPD-coding region.
XX
SQ Sequence 4430 BP; 1333 A; 871 C; 955 G; 1271 T; 0 other;

Query Match 18.0%; Score 154.8; DB 16; Length 4430;
Best Local Similarity 88.4%; Pred. No. 1.5e-25;
Matches 168; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

```
Qy     246 CAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGG 305
           | | | ||||     || || |||| || ||||||||||||||| |||| ||||||||||
Db    1784 ccaaaactttctctcttgataaggatatagtgacaaatgtttcacccagaatcatccggg 1843

Qy     306 GAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCA 365
           | ||||||||||||||||||||||||||||||||||||||| || ||||||||||||||||
Db    1844 ggaccacctctggccccatgtatggccctggacaaagctctttcctgaatattgagctca 1903

Qy     366 TCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCC 425
           ||||||| |||||||||||||||||||||||||||||||||||||||| || ||||| |
Db    1904 tcagtgaaaaaacggctgcatattggtgtcaaagtgtcactgaactaaaagccgactttc 1963

Qy     426 CAGACAACaT 435
           ||||||| ||
Db    1964 cagacaatat 1973
```

RESULT 8
AAF58252
ID AAF58252 standard; DNA; 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.

XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 17.7%; Score 152.6; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 3.4e-25;
Matches 8; Conservative 512; Mismatches 271; Indels 0; Gaps 0;

```
Qy       9 AAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGTAATTTATC 68
           :: :::::::::::: ::::::   : :: :::: ::   ::: :: :: ::::::::
Db       2 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 61

Qy      69 TGGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTTCAATTCTA 128
           :  :  ::: ::::: ::::::::::  ::::::::::  :: :: ::::: :: :::: ::
Db      62 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwww 121

Qy     129 ATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATAAAATATTG 188
           ::::::: :::::::::::::  : :  :::::: : :  :   :  : :::::::::
Db     122 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 181

Qy     189 TCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGGAGACTCAA 248
           : ::::  :::: :| : :::::::: ::::::  :::: ::: : ::::  : : : ::
Db     182 wwwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 241

Qy     249 TATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAA 308
           ::: :::: : ::: :: :  : ::: : : :::: :::     ::::: ::|     ::
Db     242 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwww 301

Qy     309 CCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCA 368
           :  : :       :: :::    :  : :::  :  ::: : :::::: :  : :: :
Db     302 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 361

Qy     369 GTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAG 428
           : : ::::    :  :::::  : : :::|  | : : :: ::::   : : ::   :
Db     362 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwww 421

Qy     429 ACAACaTAAGTGTGATAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTA 488
           : :: :::: : : :::::::: ::::: :: : ::::   :::: ::  : ::: ::::
Db     422 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 481

Qy     489 TTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATCTG 548
           :::::: ::  : :: :::   :::::::::::::: : :::::::::: ::  ::: :
Db     482 wwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 541

Qy     549 CCAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAA 608
             : ::  :::  : ::  ::: ::: :::::::: ::: ::::  : ::|:: : ::::
Db     542 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwww 601

Qy     609 AAACCCACACAATGTGAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATC 668
           :::   : : ::: : ::  : : :::::::::   :   :: ::: ::  :: :: :::
Db     602 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 661

Qy     669 TACATAAGCAGAAGGAAGAGAAATGAAATACTCATTTTATTGAGTTGGCCCCCACTGTAT 728
           :: :::: : ::  :: : :::: ::: : : :::::::: : ::       : : :::
Db     662 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwww 721

Qy     729 GTGGCTGGTATTTATGAAGGTGATGACCCAGGAAGAAATTGTAAACTATAAACCACTCCA 788
           :   :  ::::::: ::  : :: :   :  :: ::::: :::: :::::::  : :  :
Db     722 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 781

Qy     789 AATATAAACCC 799
           ::::::  |||
Db     782 wwwwwwgcccc 792
```

RESULT 9

```
AAF58254
ID   AAF58254 standard; DNA; 936 BP.
XX
AC   AAF58254;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1875.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface  -
XX
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;
```

Query Match 17.7%; Score 152.6; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 3.4e-25;
Matches 8; Conservative 512; Mismatches 271; Indels 0; Gaps 0;

```
Qy       9 AAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGTAATTTATC 68
Db       2 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 61

Qy      69 TGGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTTCAATTCTA 128
Db      62 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwww 121

Qy     129 ATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATAAAATATTG 188
Db     122 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 181

Qy     189 TCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGGAGACTCAA 248
Db     182 wwwwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 241

Qy     249 TATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAA 308
Db     242 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwww 301

Qy     309 CCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCA 368
Db     302 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 361

Qy     369 GTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAG 428
Db     362 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwww 421

Qy     429 ACAACaTAAGTGTGATAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTA 488
Db     422 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 481

Qy     489 TTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATCTG 548
Db     482 wwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 541

Qy     549 CCAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAA 608
Db     542 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwww 601

Qy     609 AAACCCACACAATGTGAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATC 668
Db     602 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 661

Qy     669 TACATAAGCAGAAGGAAGAGAAATGAAATACTCATTTTATTGAGTTGGCCCCCACTGTAT 728
Db     662 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwww 721

Qy     729 GTGGCTGGTATTTATGAAGGTGATGACCCAGGAAGAAATTGTAAACTATAAACCACTCCA 788
Db     722 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 781

Qy     789 AATATAAACCC 799
Db     782 wwwwwwgcccc 792
```

```
RESULT  10
AAF58257
ID   AAF58257 standard; DNA; 936 BP.
XX
AC   AAF58257;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1954.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface  -
XX
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
```

```
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;


  Query Match             17.7%;  Score 152.6;  DB 22;  Length 936;
  Best Local Similarity   1.0%;  Pred. No. 3.4e-25;
  Matches    8;  Conservative  512;  Mismatches  271;  Indels    0;  Gaps    0;

Qy       9 AAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGTAATTTATC 68
           :: :::::::::::: :::::::   : :: ::::  ::   ::: :: :: ::::::::
Db       2 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 61

Qy      69 TGGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTTCAATTCTA 128
           :  :  ::: ::::: ::::::::::  :::::::::  :: :: ::::: :: :::: ::
Db      62 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwww 121

Qy     129 ATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATAAAATATTG 188
           ::::::: :::::::::::::  : :  :::::: : :  :   :  : ::::::::::
Db     122 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 181

Qy     189 TCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGGAGACTCAA 248
           : ::::  :::: :| : :::::::: ::::::  :::: ::: : ::::  : : : ::
Db     182 wwwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 241

Qy     249 TATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAA 308
           ::: :::: : ::: :: :  : ::: : : :::: :::      ::::: ::|     ::
Db     242 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwww 301

Qy     309 CCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCA 368
             : : :      :: :::     :  : :::  :  ::: : :::::: :  : :: :
Db     302 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 361

Qy     369 GTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAG 428
           : : ::::    :  :::::  : : :::|  | : : :: ::::   : : ::   :
Db     362 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwww 421

Qy     429 ACAACaTAAGTGTGATAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTA 488
           : :: :::: : : :::::::: ::::: :: : ::::   :::: ::  : ::: ::::
Db     422 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 481

Qy     489 TTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATCTG 548
           :::::: ::  : : :: :::  :::::::::::::: : :::::::::: ::  ::: :
Db     482 wwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 541

Qy     549 CCAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAA 608
           : ::  :::  : ::  ::: ::: :::::::: ::: ::::  : ::|:: : ::::
Db     542 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwww 601

Qy     609 AAACCCACACAATGTGAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATC 668
           :::   : : ::: : ::  : : :::::::   :   :: ::: ::  :: :: :::
Db     602 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 661

Qy     669 TACATAAGCAGAAGGAAGAGAAATGAAATACTCATTTTATTGAGTTGGCCCCCACTGTAT 728
           :: ::::  : ::  :: : :::: ::: : : :::::::: : ::       : : :::
Db     662 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwww 721

Qy     729 GTGGCTGGTATTTATGAAGGTGATGACCCAGGAAGAAATTGTAAACTATAAACCACTCCA 788
           :   :  ::::::: ::  : :: :   :  :: ::::: :::: ::::::  : :  :
Db     722 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 781

Qy     789 AATATAAACCC 799
           ::::::  |||
Db     782 wwwwwwgcccc 792


RESULT  11
AAF58259
ID   AAF58259 standard; DNA; 936 BP.
XX
AC   AAF58259;
XX
```

```
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D2004.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface  -
XX
PS   Example 6; Page 128; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;


  Query Match             17.7%;  Score 152.6;  DB 22;  Length 936;
  Best Local Similarity   1.0%;  Pred. No. 3.4e-25;
  Matches    8;  Conservative  512;  Mismatches  271;  Indels    0;  Gaps    0;

Qy       9 AAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGTAATTTATC 68
           :: :::::::::::: :::::::   : :: ::::  ::   ::: :: :: ::::::::
Db       2 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 61

Qy      69 TGGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTTCAATTCTA 128
           :  :  ::: ::::: ::::::::::  :::::::::  :: :: ::::: :: :::: ::
Db      62 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwww 121

Qy     129 ATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATAAAATATTG 188
           ::::::: :::::::::::::  : :  :::::: : :  :   :  : ::::::::::
Db     122 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 181

Qy     189 TCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGGAGACTCAA 248
           : ::::  :::: :| : :::::::: ::::::  :::: ::: : ::::  : : : ::
Db     182 wwwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 241

Qy     249 TATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAA 308
           ::: :::: : ::: :: :  : ::: : : :::: :::      ::::: ::|     ::
Db     242 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwww 301

Qy     309 CCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCA 368
             : : :      :: :::     :  : :::  :  ::: : :::::: :  : :: :
Db     302 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 361

Qy     369 GTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAG 428
           : : ::::    :  :::::  : : :::|  | : : :: ::::   : : ::   :
Db     362 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwww 421

Qy     429 ACAACaTAAGTGTGATAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTA 488
```

```
                : :: :::: : : :::::::: ::::: :: : ::::   :::: ::  : ::: ::::
Db       422 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 481

Qy       489 TTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATCTG 548
             :::::: ::  : :: :::   ::::::::::::: : ::::::::: ::  ::: :
Db       482 wwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 541

Qy       549 CCAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAA 608
               : ::  :::  : ::  ::: ::: :::::::: ::: ::::  : ::|:: : ::::
Db       542 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwww 601

Qy       609 AAACCCACACAATGTGAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATC 668
             :::    : : ::: : ::  : : ::::::::   :    :: ::: ::  :: :: :::
Db       602 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 661

Qy       669 TACATAAGCAGAAGGAAGAGAAATGAAATACTCATTTTATTGAGTTGGCCCCCACTGTAT 728
             :: ::::  : ::  :: : :::: ::: : : :::::::: : ::         : : :::
Db       662 wwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 721

Qy       729 GTGGCTGGTATTTATGAAGGTGATGACCCAGGAAGAAATTGTAAACTATAAACCACTCCA 788
               :    :  ::::::: ::  : :: :   :  :: ::::: :::: ::::::  : :  :
Db       722 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 781

Qy       789 AATATAAACCC 799
             ::::::  |||
Db       782 wwwwwwgcccc 792
```

RESULT 12
AAF58262
ID   AAF58262 standard; DNA; 936 BP.
XX
AC   AAF58262;
XX
DT   24-APR-2001 (first entry)
XX
DE   Oligonucleotide D2007.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999; 99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface -
XX
PS   Example 6; Page 128; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

```
  Query Match             17.7%;  Score 152.6;  DB 22;  Length 936;
  Best Local Similarity   1.0%;  Pred. No. 3.4e-25;
  Matches    8;  Conservative  512;  Mismatches  271;  Indels    0;  Gaps    0;

Qy         9 AAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGTAATTTATC 68
             :: :::::::::::: ::::::    : :: ::::  ::    ::: :: :: ::::::::
Db         2 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 61

Qy        69 TGGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTTCAATTCTA 128
               :   :  ::: ::::: :::::::::  :::::::::  :: :: ::::: :: :::: ::
Db        62 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwww 121

Qy       129 ATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATAAAATATTG 188
             ::::::: ::::::::::::   : :   :::::: : :    :    :  : :::::::::
Db       122 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 181

Qy       189 TCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGGAGACTCAA 248
             : ::::   :::: :| : :::::::: ::::::  :::: ::: : ::::   : : : ::
Db       182 wwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 241

Qy       249 TATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAA 308
             ::: :::: : ::: :: :   : ::: : : :::: :::      ::::: ::|      ::
Db       242 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwww 301

Qy       309 CCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCA 368
                :   : :        :: :::      :   : :::   :   ::: : :::::: :   : :: :
Db       302 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 361

Qy       369 GTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAG 428
               : : ::::     :   :::::   : : :::|   | : : :: ::::    : : ::    :
Db       362 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwww 421

Qy       429 ACAACaTAAGTGTGATAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTA 488
             :::: :::: : : :::::::: ::::: :: : ::::    :::: ::   : ::: ::::
Db       422 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 481

Qy       489 TTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATCTG 548
             :::::: ::  : :: :::   ::::::::::::: : ::::::::: ::  ::: :
Db       482 wwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 541

Qy       549 CCAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAA 608
               : ::  :::  : ::  ::: ::: :::::::: ::: ::::  : ::|:: : ::::
Db       542 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwww 601

Qy       609 AAACCCACACAATGTGAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATC 668
             :::    : : ::: : ::  : : ::::::::   :    :: ::: ::  :: :: :::
Db       602 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 661

Qy       669 TACATAAGCAGAAGGAAGAGAAATGAAATACTCATTTTATTGAGTTGGCCCCCACTGTAT 728
             :: ::::  : ::  :: : :::: ::: : : :::::::: : ::         : : :::
Db       662 wwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 721

Qy       729 GTGGCTGGTATTTATGAAGGTGATGACCCAGGAAGAAATTGTAAACTATAAACCACTCCA 788
               :    :  ::::::: ::  : :: :   :  :: ::::: :::: ::::::  : :  :
Db       722 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 781

Qy       789 AATATAAACCC 799
             ::::::  |||
Db       782 wwwwwwgcccc 792
```

RESULT 13
AAF58255
ID   AAF58255 standard; DNA; 938 BP.
XX
AC   AAF58255;
XX
DT   24-APR-2001 (first entry)
XX
DE   Oligonucleotide D1876.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;

```
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface
XX
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

  Query Match             17.7%;  Score 152.6;  DB 22;  Length 938;
  Best Local Similarity   1.0%;  Pred. No. 3.4e-25;
  Matches    8;  Conservative  512;  Mismatches  271;  Indels    0;  Gaps    0;

Qy       9 AAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGTAATTTATC 68
Db       2 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 61

Qy      69 TGGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTTCAATTCTA 128
Db      62 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwww 121

Qy     129 ATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATAAAATATTG 188
Db     122 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 181

Qy     189 TCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGGAGACTCAA 248
Db     182 wwwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 241

Qy     249 TATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAA 308
Db     242 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwww 301

Qy     309 CCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCA 368
Db     302 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 361

Qy     369 GTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAG 428
Db     362 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwww 421

Qy     429 ACAACaTAAGTGTGATAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTA 488
Db     422 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 481

Qy     489 TTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATCTG 548
Db     482 wwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 541

Qy     549 CCAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAA 608
Db     542 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwww 601

Qy     609 AAACCCACACAATGTGAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATC 668
Db     602 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 661

Qy     669 TACATAAGCAGAAGGAAGAGAAATGAAATACTCATTTTATTGAGTTGGCCCCCACTGTAT 728
Db     662 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwww 721

Qy     729 GTGGCTGGTATTTATGAAGGTGATGACCCAGGAAGAAATTGTAAACTATAAACCACTCCA 788
Db     722 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 781

Qy     789 AATATAAACCC 799
Db     782 wwwwwwgcccc 792


RESULT  14
AAT14078
ID   AAT14078 standard; cDNA; 4447 BP.
XX
AC   AAT14078;
XX
DT   11-JUN-1996  (first entry)
XX
DE   Pig dihydropyrimidine dehydrogenase cDNA.
XX
KW   Dihydropyrimidine dehydrogenase; DPD; DPYD gene; probe; deficiency;
KW   5-fluorouracil; cytostatic; cancer; selectable marker; ss.
XX
OS   Sus scrofa.
XX
FH   Key             Location/Qualifiers
FT   CDS             88..3165
FT                   /*tag=  a
XX
PN   WO9608568-A2.
XX
PD   21-MAR-1996.
XX
PF   07-SEP-1995;   95WO-US12016.
XX
PR   12-SEP-1994;   94US-0304309.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Fernandez-Salguero P,  Gonzalez FJ;
XX
DR   WPI; 1996-179943/18.
DR   P-PSDB; AAR91421.
XX
PT   Pig and human di:hydro:pyrimidine dehydrogenase (DPD) genes and
PT   probes  - useful for detection of DPD deficiencies and
PT   identification of humans at risk of toxic reaction to
PT   5-fluoro:uracil anti-cancer treatment
XX
PS   Claim 1; Page 52-56; 78pp; English.
XX
CC   Pig dihydropyrmidine dehydrogenase (DPD) cDNA (AAT14078) was
CC   isolated from a pig liver cDNA library by screening with a
CC   polyclonal antibody raised against pig DPD, and rescreening
CC   using an isolated partial cDNA.  The DPD cDNA can be inserted
CC   into a vector and used for prodn. of active DPD (AAR91421) in
CC   transformed host cells, e.g. Escherichia coli, or used as a
CC   selectable marker functional in prokaryotic and eukaryotic cells.
CC   It can also be utilised as a probe to detect DPD deficiency in
CC   an individual, i.e. to identify risk of toxic reaction to
```

```
CC   4-fluorouracil.
XX
SQ   Sequence 4447 BP; 1285 A; 928 C; 1000 G; 1234 T; 0 other;

  Query Match            17.6%;  Score 151.6;  DB 17;  Length 4447;
  Best Local Similarity  87.4%;  Pred. No. 7.9e-25;
  Matches  166;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy     246 CAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGG 305
           | | | ||||     || || ||||||| |||||||||| || |||| |||| ||||||
Db    1805 ccaaaactttctctcttgataaggacatagtgacaaatgtctcacccagaatcgtccggg 1864

Qy     306 GAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCA 365
           | || ||||||||||||||||| |||||||||||||||||||| ||||||||||||||||
Db    1865 ggactacctctggccccatgtacggccctggacaaagctccttcctgaatattgagctca 1924

Qy     366 TCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCC 425
           ||||||| ||||| |||||||||||||||||||||||||||||||||| |||||||| |
Db    1925 tcagtgaaaaaacagctgcatattggtgtcaaagtgtcactgaactaaaagctgactttc 1984

Qy     426 CAGACAACaT 435
           ||||||| ||
Db    1985 cagacaatat 1994


RESULT  15
AAF58252/c
ID   AAF58252 standard; DNA; 936 BP.
XX
AC   AAF58252;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1835.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface  -
XX
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

  Query Match            17.3%;  Score 148.6;  DB 22;  Length 936;
  Best Local Similarity   0.5%;  Pred. No. 2.7e-24;
  Matches    4;  Conservative  511;  Mismatches  270;  Indels    0;  Gaps    0;

Qy      12 ATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGTAATTTATCTGG 71
           :::::::::::: ::::::   : :: ::::  ::   ::: :: :: :::::::: :
Db     786 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 727

Qy      72 AGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTTCAATTCTAATT 131
           :  ::: ::::: :::::::::  :::::::::  :  :: ::::: :: :::: :::::
Db     726 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWW 667

Qy     132 TAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATAAAATATTGTCA 191
           :::: ::::::::::::::  : :  :::::: : :  :   :  : :::::::::: : :
Db     666 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 607

Qy     192 TATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGGAGACTCAATAT 251
           :::  :::: :  : :::::::: ::::::  :::: ::: : ::::  : : : :::::
Db     606 WWWWWWWWWWWWWWCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 547

Qy     252 CTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAACCA 311
           :::: : ::: :: :  : ::: : : :::: :::     ::::: ::   |  ::  :
Db     546 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWW 487

Qy     312 CCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCAGTG 371
             : :      :: :::     :  : :::  :  ::: : :::::: :  : :: : :
Db     486 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 427

Qy     372 AGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAGACA 431
           : ::::    :  ::::::  : : ::: : | |   :: ::::   : : ::   : : :
Db     426 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWTAAGCWWWWWWWWWWWWWWWWWW 367

Qy     432 ACaTAAGTGTGATAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTATTT 491
           : :::: : : :::::::: ::::: :: : ::::   :::: ::  : ::: :::::::
Db     366 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 307

Qy     492 AAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATCTGCCA 551
           ::: ::  :::: :::   :::::::::::::: : :::::::::: ::  ::: :   :
Db     306 WWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 247

Qy     552 GTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAAAAA 611
           ::  :::  : ::  ::: ::: :::::::: ::: :::  : :: ::|: :::::::
Db     246 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWCWWWWWWWW 187

Qy     612 CCCACACAATGTGAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATCTAC 671
             : : ::: : ::  : : :::::::   :   :: ::: ::  :: :: ::: ::
Db     186 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 127

Qy     672 ATAAGCAGAAGGAAGAGAAATGAAATACTCATTTTATTGAGTTGGCCCCCACTGTATGTG 731
           ::::  : ::  :: : :::: :::::   :::::::: : ::       : : ::: :
Db     126 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWW 67

Qy     732 GCTGGTATTTATGAAGGTGATGACCCAGGAAGAAATTGTAAACTATAAACCACTCCAAAT 791
           :  ::::::: ::  : :: :   :  :: ::::: :::: ::::::  : :  ::::
Db      66 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 7

Qy     792 ATAAA 796
           :::::
Db       6 WWWWW 2


Search completed: November 20, 2001, 16:49:18
Job time: 4632 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2001, 15:07:46 ; Search time 1366.05 Seconds
(without alignments)
5957.986 Million cell updates/sec

Title: US-09-308-080-1
Perfect score: 861
Sequence: 1 TGTTAATGAAGATAAATATT.....AGTGGGAAATAATTTATTAA 861

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*
190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	171.2	19.9	938	172	BG024298	BG024298 602274460
c 2	90.8	10.5	853	220	CNS02EEI	AL193635 Tetraodon
3	87.8	10.2	621	110	AW018806	AW018806 fd60d03.y
4	71.4	8.3	1207	219	CNS00175	AL058798 Drosophil
5	64	7.4	562	165	BE288474	BE288474 601095732
6	58	6.7	1101	219	CNS0039G	AL063921 Drosophil
7	55	6.4	500	107	AU086287	AU086287 AU086287
8	54.8	6.4	483	151	BF653737	BF653737 277669 MA
c 9	52.6	6.1	1101	219	CNS00EVL	AL069706 Drosophil
10	52	6.0	912	245	AZ527678	AZ527678 ENTBN23TR
11	52	6.0	1101	219	CNS00EO7	AL069440 Drosophil
c 12	51.8	6.0	853	245	AZ548198	AZ548198 ENTFW05TF
c 13	51.8	6.0	868	248	AZ686179	AZ686179 ENTJI41TR
c 14	51.8	6.0	881	247	AZ681140	AZ681140 ENTMC34TR
15	51.8	6.0	899	245	AZ542627	AZ542627 ENTCX21TR
16	51.4	6.0	898	248	AZ686287	AZ686287 ENTKG36TR
17	51.4	6.0	1101	219	CNS00Z0C	AL097062 Drosophil
18	51.2	5.9	987	219	CNS014PQ	AL104456 Drosophil
c 19	51.2	5.9	1025	219	CNS014J2	AL104216 Drosophil
c 20	51	5.9	1101	219	CNS0039G	AL063921 Drosophil
21	50.2	5.8	1101	219	CNS017V6	AL108540 Drosophil
22	50	5.8	632	107	AU060843	AU060843 AU060843
c 23	49.8	5.8	947	247	AZ682793	AZ682793 ENTKY77TF
24	49.8	5.8	1101	219	CNS016LI	AL106896 Drosophil
25	49.6	5.8	867	219	CNS00CX5	AL060052 Drosophil
c 26	49.4	5.7	872	245	AZ535541	AZ535541 ENTBP32TR
c 27	49.4	5.7	1101	219	CNS003DQ	AL064580 Drosophil
c 28	49	5.7	896	245	AZ534072	AZ534072 ENTBR33TR
c 29	48.8	5.7	816	220	CNS02KJV	AL201604 Tetraodon
c 30	48.6	5.6	872	245	AZ527849	AZ527849 ENTCF51TR
31	48.4	5.6	329	105	AL513719	AL513719 AL513719
32	48.4	5.6	911	248	AZ686515	AZ686515 ENTIU42TR
c 33	48.2	5.6	822	219	CNS009CW	AL053618 Drosophil
34	48.2	5.6	898	247	AZ674122	AZ674122 ENTIB47TF
c 35	48.2	5.6	987	219	CNS014PQ	AL104456 Drosophil
36	48.2	5.6	1007	219	CNS01460	AL103746 Drosophil
c 37	48	5.6	892	245	AZ549872	AZ549872 ENTFG30TR
c 38	47.4	5.5	872	248	AZ683520	AZ683520 ENTIO43TR
c 39	47.4	5.5	878	219	CNS0187R	AL108993 Drosophil
40	47.4	5.5	959	219	CNS00655	AL062806 Drosophil
41	47.4	5.5	1190	220	CNS02ON7	AL206908 Tetraodon
c 42	47.2	5.5	909	248	AZ690408	AZ690408 ENTLW08TR
43	47.2	5.5	1032	220	CNS02OIP	AL206746 Tetraodon
c 44	47.2	5.5	1048	248	AZ683764	AZ683764 ENTIL96TR
45	47.2	5.5	1101	219	CNS00EVL	AL069706 Drosophil

ALIGNMENTS

```
RESULT   1
BG024298
LOCUS       BG024298      938 bp    mRNA            EST       24-JAN-2001
DEFINITION  602274460F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4362302 5',
            mRNA sequence.
ACCESSION   BG024298
VERSION     BG024298.1  GI:12409730
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 938)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis Staudt, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10006  row: h  column: 15
            High quality sequence stop: 714.
FEATURES             Location/Qualifiers
     source          1. .938
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4362302"
                     /clone_lib="NIH_MGC_85"
                     /tissue_type="lymphoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.867 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      222 a    234 c    253 g    229 t
ORIGIN


  Query Match             19.9%;  Score 171.2;  DB 172;  Length 938;
  Best Local Similarity   93.2%;  Pred. No. 4.4e-29;
  Matches  179;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy     244 CTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCG 303
           | | | | ||||     || || |||||||||||||||||||||||||| |||||||||
Db     260 CACCAAAACTTTCTCTCTTGATAAGGACATTGTGACAAATGTTTCCCCCAGAATCATCCG 319

Qy     304 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     320 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 379

Qy     364 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 423
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     380 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 439

Qy     424 CCCAGACAACaT 435
           ||||||||||||
Db     440 TCCAGACAACAT 451


RESULT   2
CNS02EEI/c
LOCUS       CNS02EEI      853 bp    DNA             GSS       13-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            262A08 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL193635
VERSION     AL193635.1  GI:7831741
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 853)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 853)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
```

```
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 853)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .853
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="262A08"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG262BA04LP1~end : T7"
BASE COUNT      252 a    189 c    136 g    269 t      7 others
ORIGIN
```

```
  Query Match             10.5%;  Score 90.8;  DB 220;  Length 853;
  Best Local Similarity   67.9%;  Pred. No. 1e-10;
  Matches  125;  Conservative    1;  Mismatches   58;  Indels    0;  Gaps    0;

Qy     258 TCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAACCACCTCTG 317
           | ||||   ||||| | || |||||||| || ||    ||  | || || ||||| || |
Db     409 TGTTTCTGCAGGACTTGGTAACAAATGTGTCTCCTCGTATTGTGCGAGGCACCACTTCAG 350

Qy     318 GCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCAGTGAGAAAA 377
           ||    |||||||||| || ||  | || || || || |||||||| || || ||||| |
Db     349 GCAATGTGTATGGCCCAGGTCAGGGGTCATTCCTCAACATTGAGCTTATTAGCGAGAAGA 290

Qy     378 CGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAGACAACaTAA 437
           |||| || ||||||||||||   ||  | |:||| |||     ||| ||  | ||  |||
Db     289 CGGCAGCTTATTGGTGTCAATCAGTGGCCGMACTTAAGAGGAGCTTTCCGAATAATGTAA 230

Qy     438 GTGT 441
           || |
Db     229 GTTT 226
```

```
RESULT   3
AW018806
LOCUS       AW018806      621 bp    mRNA            EST       13-SEP-1999
DEFINITION  fd60d03.y1 Zebrafish WashU MPIMG EST Danio rerio cDNA 5' similar to
            SW:DPYD_HUMAN Q12882 DIHYDROPYRIMIDINE DEHYDROGENASE [NADP+]
            PRECURSOR ;, mRNA sequence.
ACCESSION   AW018806
VERSION     AW018806.1  GI:5872336
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 621)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 491.
FEATURES             Location/Qualifiers
     source          1. .621
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone_lib="Zebrafish WashU MPIMG EST"
                     /sex="mixed"
                     /tissue_type="26 somite embryos, adult livers, shield
                     stage embryos"
                     /lab_host="XL1-blue MRF"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
                     strand cDNA was primed with a Not I - oligo(dT)15 primer
                     [5'pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTT3']
                     double-stranded cDNA was ligated to Sal I adaptors (BRL),
                     digested with Not I and cloned into the Not I and Sal I
                     sites of the pSPORT1 vector (BRL). Library was constructed
                     by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
                     Institut fuer Molekulare Genetik,Berlin). cDNAs for EST
                     analysis were selected following oligonucleotide
                     hybridization fingerprinting of arrayed clones from
                     zebrafish late somitogenesis (26 ss), adult liver or
                     embryonic shield stage (5.6 h) libraries.  Fingerprint
                     data were used to computationally cluster cDNAs, and a
                     single cDNA from each cluster was chosen for sequencing.
                     In some cases multiple members of the same cluster were
                     sequenced to assess clustering parameters or single clones
                     were sequenced additional times to assess quality
                     control."
BASE COUNT      151 a    169 c    160 g    141 t
ORIGIN
```

```
  Query Match             10.2%;  Score 87.8;  DB 110;  Length 621;
  Best Local Similarity   71.2%;  Pred. No. 4.9e-10;
  Matches  116;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy     275 GTGACAAATGTTTCCCCCATAATCATCCGGGGAACCACCTCTGGCCCCATGTATGGCCCT 334
           ||||| || || || ||   |||| | || || || ||||| || | ||| |  || |||
Db       1 GTGACCAACGTGTCACCGCGAATCGTGCGTGGGACAACCTCAGGTCACATATTCGGTCCT 60

Qy     335 GGACAAAGCTCCTTTCTGAATATTGAGCTCATCAGTGAGAAAACGGCTGCATATTGGTGT 394
           || ||  |||| || || || || |||||||| ||||| |||||| || || || |||||
Db      61 GGTCAGGGCTCATTCCTCAACATCGAGCTCATTAGTGAAAAAACTGCCGCTTACTGGTGC 120

Qy     395 CAAAGTGTCACTGAACTAAAGGCTGACTTCCCAGACAACaTAA 437
           | || ||  ||||||| || || ||||||||  | || || |
Db     121 AAGAGCGTGGCTGAACTCAAAGCCGACTTCCCCAAAAATATTA 163
```

```
RESULT   4
CNS00175
LOCUS       CNS00175     1207 bp    DNA             GSS       04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR03G12 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL058798
VERSION     AL058798.1  GI:4939103
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1207)
```

```
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1207
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR03G12"
                     /note="end : TET3"
BASE COUNT      250 a    298 c    328 g    233 t     98 others
ORIGIN


  Query Match             8.3%;  Score 71.4;  DB 219;  Length 1207;
  Best Local Similarity   63.9%;  Pred. No. 2.8e-06;
  Matches  108;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy    267 AGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAACCACCTCTGGCCCCATGT 326
          ||||  | || || ||||| || ||    ||| || |||| ||||| || |||  || ||
Db    399 AGGATCTGGTCACGAATGTCTCGCCGCGCATCGTCAGGGGCACCACGTCGGGCTACAAGT 458

Qy    327 ATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCAGTGAGAAAACGGCTGCAT 386
          |||| ||    ||  |||  || ||||| || || || |||   |||||   ||| |  |
Db    459 ATGGACCGCAGCAGGGCTGTTTCCTGAACATCGAACTAATCTCGGAGAAGCGGGCCGAGT 518

Qy    387 ATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAGACAACaT 435
          | |||    ||    |    ||||| |||  |||||||||| || || ||
Db    519 ACTGGCTGAAATCGATTGGAGAACTGAAGCGTGACTTCCCCGAGAAGAT 567


RESULT   5
BE288474
LOCUS       BE288474      562 bp    mRNA            EST       26-OCT-2000
DEFINITION  601095732F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3490423 5',
            mRNA sequence.
ACCESSION   BE288474
VERSION     BE288474.1  GI:9168194
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 562)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8532  row: p  column: 08
            High quality sequence stop: 562.
FEATURES             Location/Qualifiers
     source          1. .562
                     /organism="Mus musculus"
                     /strain="C57/B6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3490423"
                     /clone_lib="NCI_CGAP_Mam5"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="7 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Robin Humphreys,
                     NIH"
BASE COUNT      137 a    124 c    166 g    135 t
ORIGIN


  Query Match             7.4%;  Score 64;  DB 165;  Length 562;
  Best Local Similarity   93.1%;  Pred. No. 0.00013;
  Matches   67;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy    364 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 423
          |||||||||||||||| |||||||||||||||| |||||||| ||||||||||||||||
Db      1 CATCAGTGAGAAAACAGCTGCATATTGGTGTCACAGTGTCACCGAACTAAAGGCTGACTT 60

Qy    424 CCCAGACAACaT 435
          ||| ||||| ||
Db     61 CCCGGACAATAT 72


RESULT   6
CNS0039G
LOCUS       CNS0039G      1101 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
```

```
                     /clone_lib="RPCI-98"
                     /clone="BACR08K10"
                     /note="end : TET3"
BASE COUNT      201 a      64 c     131 g     202 t     503 others
ORIGIN
```

Query Match 6.7%; Score 58; DB 219; Length 1101;
Best Local Similarity 20.4%; Pred. No. 0.0032;
Matches 116; Conservative 216; Mismatches 232; Indels 6; Gaps 1;

```
Qy      70 GGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTTCAATTCTAA 129
           || |:  | |:|:: ::| ||||   :   :|     |  ::  : :     | |:  ||
Db     391 GGGGMATATAWAWWWWWTTTTTTTTAWAAAWAAAATAATTWWAAWAWAAAAAATTWWAAA 450

Qy     130 TTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATAAAATATTGT 189
           : ||   :||::   |:|:       ||| || :            |:| |::|| :
Db     451 AWAAAAWAWTAWWTTTAWTWAAAAAAAAAAAATTWTTTTTTTWTTTAWTTWATAWWTTWW 510

Qy     190 CATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGGAGACTCAAT 249
             ||   |||  |  | |:|| :|| : || |  |::| :|| | :||:  | |: :||:
Db     511 WTTAAAWAAAAAAAAAAAAWAAAWAAAWATAAATWTWWTWWTTYTTWAAWATAAAMCMAAW 570

Qy     250 ATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAAC 309
           ::|:| :|:::: :     :::   :|:|: |::: ::   ::: :| :    |:
Db     571 YYHTYTTYTYHYYTYWTYTMTWHYHTMYTHAWAHTTWYHWYHTYAMWHWMTWHWTMWAWW 630

Qy     310 CACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCAG 369
           ::  :    ::::   : :  :: :  : :||  :  :|: | | :: | : :::::||
Db     631 HWTTYTAAYYYYYTCMYYYHYMHWHHAHAHAAAWWTTHTWWTHAYHWATYHYYYYMYCAM 690

Qy     370 TGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAGA 429
             :  :  :  :: ::: |   | |:::  : ::::: :|:::       || | :: :
Db     691 MCMCTHTCHHCYYYYHHYTAHHTHTHHWYAHYYMWYYMWAYYWM------YCTACTYHYH 744

Qy     430 CAACaTAAGTGTGATAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTAT 489
           :::: ::| : |   :|:|::::::|::|| | :| :   :: : ||    : : :::
Db     745 HHHHYHWAYHTTWYAWAHAMWMWHHAHYAAAAAWAAWATTHHYHHTTHYMHHTYMYHYYM 804

Qy     490 TTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATCTGC 549
           :|   : |:: :::::||  |:|:|::: :|: :: : |::::|||:: ::: : :|:
Db     805 YTCCYMCTYHCWHYYHTAYTCWTWTHHWMWTWTHWYHHTWWHHTTTHWAWWHTHTWCWWW 864

Qy     550 CAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAAA 609
           : ||  : | |:    :::: ::: :::::| : ::: :: : :::::   ::::|: :
Db     865 WHATTWTWATHCWACMTMHWHHWMHMHHHHMACHAHHHTHMCMCHHHMHCTCHHHHTMYH 924

Qy     610 AACCCACACAATGTGAAGCTCTGTTATAAA 639
           : |: : ::::: : ::  |   |  |:::
Db     925 MTCHWMWHMHWHWHHWWMATWMTTMTTMMM 954

```
RESULT 7
AU086287
LOCUS AU086287 500 bp mRNA EST 27-JAN-2001
DEFINITION AU086287 Sugano Malaria cDNA library Plasmodium falciparum cDNA clone XPFn2249, mRNA sequence.
ACCESSION AU086287
VERSION AU086287.1 GI:12388428
KEYWORDS EST.
SOURCE malaria parasite P. falciparum.
  ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 500)
  AUTHORS Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.
  TITLE FULL-malaria: a database for a full-length enriched cDNA library from human malaria parasite, Plasmodium falciparum
  JOURNAL Nucleic Acids Res. 29 (1), 70-71 (2001)
  MEDLINE 20574754
COMMENT Contact: Junichi Watanabe
Institute of Medical Science
The University of Tokyo, Department of Parasitology
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Tel: 81-3-5449-5378
Fax: 81-3-5449-5410
Email: jwatanab@manage.ims.u-tokyo.ac.jp
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano ,S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).
FEATURES Location/Qualifiers
     source 1. .500
/organism="Plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/clone="XPFn2249"
/clone_lib="Sugano Malaria cDNA library"
/dev_stage="erythrocytic stage"
BASE COUNT 214 a 25 c 24 g 231 t 6 others
ORIGIN
```

Query Match 6.4%; Score 55; DB 107; Length 500;
Best Local Similarity 49.3%; Pred. No. 0.015;
Matches 136; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

```
Qy       7 TGAAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGTAATTTA 66
           | || || | | |||||  |||||       | |||         | || | |  || ||
Db     202 TTAATATTATTTTTTTTCATTTTTGTAATATATAATATATATTTGACTACATTTATTTAT 261

Qy      67 TCTGGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTTCAATTC 126
           | |    || ||| ||| |||||||||   ||||    | |  |  | || ||||||
Db     262 TATATTTCTGACATATAATTTATTTTATTATTTTTAAAGAATTTNNTATAAGTTCAAAAT 321

Qy     127 TAATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATAAAATAT 186
           |||  |||  ||||  |  ||   | |  | | |        |   |  | ||| || ||
Db     322 AAATAAAATCCATATATATTATATATATAAATAAAACATTTTATAAATATTATATAAAAT 381

Qy     187 TGTCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGGAGACTC 246
           | |||  | |    || ||||  |   |||||  |  |  | |  | |        |
Db     382 CTCCTTATAAACAAATACAAATAAATATATTTATATATAATTATAGATATATTTTTGTGC 441

Qy     247 AATATCTTTACTCTTTCATGAGGACATTGTGACAAA 282
           | | | |||  | |||  |        || | | |||
Db     442 ATTTTTTTTNGTTTTTANTTTTTNATTTATAATAAA 477

```
RESULT 8
BF653737
LOCUS BF653737 483 bp mRNA EST 20-DEC-2000
DEFINITION 277669 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BF653737
VERSION BF653737.1 GI:11918869
KEYWORDS EST.
SOURCE cow.
  ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 483)
  AUTHORS Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid ,W.W. and Keele,J.W.
  TITLE Design and use of four pooled tissue normalized cDNA libraries for EST discovery in cattle
  JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.
PCR PRimers
FORWARD: AGGAAACAGCTATGACCAT
```

```
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 70  row: B  column: 4
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES              Location/Qualifiers
     source           1. .483
                      /organism="Bos taurus"
                      /db_xref="taxon:9913"
                      /clone_lib="MARC 3BOV"
                      /tissue_type="pooled"
                      /lab_host="DH10B"
                      /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                      Library made from pooled tissue from marrow, alveolar
                      macrophage, ovary, fetal semitendonosus muscle, and fetal
                      longissimus muscle."
BASE COUNT      128 a    105 c    138 g    112 t
ORIGIN


  Query Match             6.4%;  Score 54.8;  DB 151;  Length 483;
  Best Local Similarity   89.7%;  Pred. No. 0.017;
  Matches   70;  Conservative    0;  Mismatches    7;  Indels    1;  Gaps    1;

Qy     359 GAGCTCATCAGTG-AGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGC 417
           | ||||||||||| | |||||||||||| |||||||||||||||||||||||||||| ||
Db       1 GCGCTCATCAGTGAAAAAAACGGCTGCGTATTGGTGTCAAAGTGTCACTGAACTAAAAGC 60

Qy     418 TGACTTCCCAGACAACaT 435
           ||||| |||||||| ||
Db      61 CGACTTTCCAGACAATAT 78


RESULT  9
CNS00EVL/c
LOCUS       CNS00EVL     1101 bp    DNA             GSS       04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .1101
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone_lib="RPCI-98"
                      /clone="BACR29B23"
                      /note="end : T7"
BASE COUNT      419 a     91 c     60 g     299 t     232 others
ORIGIN


  Query Match             6.1%;  Score 52.6;  DB 219;  Length 1101;
  Best Local Similarity   35.7%;  Pred. No. 0.055;
  Matches   99;  Conservative   51;  Mismatches  127;  Indels    0;  Gaps    0;

Qy       1 TGTTAATGAAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGT 60
           | |: || |  :| :||:||||| ||||||   :  | |:||  ||       |: :| |
Db     723 TATWTATWATTWTTWATWTTTTTTTTTTTTTTTWTATTTWAAATTATTTAATTAWWWAAT 664

Qy      61 AATTTATCTGGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTT 120
           :|:|||| |     : |   | | | |     |  || |:|      | |  ||||| ||
Db     663 WAWTTATTTAATAAWTAATTAAATTATTAAAAATTTTATWAAAAAATATTTTTTTATATT 604

Qy     121 CAATTCTAATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATA 180
           | :    | ||:  ||   ||::  :|    |   |||:| | :      |    :|
Db     603 AATWAWATTTATAWATTAATATAWWTTWTAAATTWTTAATWTATTWWTTTAAATTATWTT 544

Qy     181 AAATATTGTCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGG 240
           | || : |:|:  :||  |  : ::::::||| ||:::  |||| : | :  ||:
Db     543 TTTTTTTAWTAWAWWTWAAAWAATWAWWWWWAAAAATWWWWTWTTTTTWATTWATAAWTT 484

Qy     241 AGACTCAATATCTTTACTCTTTCATGAGGACATTGTG 277
           | :|:|| :::  : : : :|||| :  :|:::
Db     483 TWTTTWWAWATTWWKKKKKKAWADTGAGARATKTDKK 447


RESULT  10
AZ527678
LOCUS       AZ527678      912 bp    DNA             GSS       03-NOV-2000
DEFINITION  ENTBN23TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   AZ527678
VERSION     AZ527678.1  GI:11079801
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 912)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 20
            High quality sequence stop: 734.
FEATURES              Location/Qualifiers
     source           1. .912
                      /organism="Entamoeba histolytica"
                      /strain="HM1:IMSS"
                      /db_xref="taxon:5759"
                      /clone_lib="Entamoeba histolytica Sheared DNA"
                      /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                      Institute for Genomic Research (TIGR), Rockville, MD.
                      Genomic DNA isolated from broth cultures of E. histolytica
                      using a  method described by Clark and Diamond (Clark,
                      C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                      method for isolate identification.  Exp. Parasitol.
                      77:450.). The DNA was mechanically sheared to give a
                      tight size distribution (~2 kb). The v + i method used for
                      the library  construction is described in detail in Smith,
```

H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999)."
BASE COUNT 466 a 51 c 81 g 314 t
ORIGIN

Query Match 6.0%; Score 52; DB 245; Length 912;
Best Local Similarity 48.3%; Pred. No. 0.075;
Matches 145; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

```
Qy    481 AATGTTTATTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCAA 540
          ||||  |||||||   |  || |  |    |  | ||||||    |  | ||   |||||
Db    179 AATGAATATTTAACATTGAAAATGTTGATTTATTTAATATTTTGTTCAACATAAAATCAA 238

Qy    541 CGAATCTGCCAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGA 600
            || |    | ||   |     ||  ||  ||   |||| ||  |   ||  |||||||
Db    239 TAAAACATATATTTTGATAATATATAAATGAAATATTAAAGAAATATACAATACTCAAGA 298

Qy    601 ACTCATAAAAACCCACACAATGTGAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAA 660
          ||| ||||||||  |   ||    | |      ||||||   |    ||   ||| | ||
Db    299 ACTAATAAAAACAAAATAAAAAGAATGAATGAATATAAAATTGATAAATAATAGAAGAAA 358

Qy    661 GAAGTATCTACATAAGCAGAAGGAAGAGAAATGAAATACTCATTTTATTGAGTTGGCCCC 720
          ||| ||   | | ||  | ||  ||   | || |||     |       |       |
Db    359 GAAATAGAAAAACAACGAAAAACAAATAATATAAAAAGAATAAAAATATATAAAAGAATA 418

Qy    721 CACTGTATGTGGCTGGTATTTATGAAGGTGATGACCCAGGAAGAAATTGTAAACTATAAA 780
            |    | || |  |     ||| ||  | || |  ||  ||||  |  ||||  |  ||
Db    419 AAAATAAAGTAGAAGAAGAATATAAAAATAATAAATCACAAAGATGTAATAAAAGAAGAA 478
```

RESULT 11
CNS00EO7
LOCUS CNS00EO7 1101 bp DNA GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069440
VERSION AL069440.1 GI:4949583
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR29P01"
/note="end : TET3"
BASE COUNT 366 a 66 c 104 g 351 t 214 others
ORIGIN

Query Match 6.0%; Score 52; DB 219; Length 1101;
Best Local Similarity 37.3%; Pred. No. 0.075;
Matches 107; Conservative 44; Mismatches 136; Indels 0; Gaps 0;

```
Qy      1 TGTTAATGAAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGT 60
            | ||| : :| ||||| |:  ||   ||||   | || || |  :|  |::| || || :
Db    561 TTTTATWWWATATAAAAAWAATTTAATTTTTWATTTTATATATTWAAWGWWAAAAAAAWW 620

Qy     61 AATTTATCTGGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTT 120
            ||  :: :  |  | | |: |: : | ||  |  : |||||||  :   : |:| |  :
Db    621 TATAAWHTWTAAATTTATAWTTWAWATTTTAAATAWATTTATTTAAWTTAWWTWTTTAAW 680

Qy    121 CAATTCTAATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATA 180
            : ||  :  |||| ::: |||:||||:     |  :||||    |  |         |
Db    681 AWTTTAAWTATTAAARWWAATTWAAAAWAAAAWTTTWAATAAWAAATAAAAATAAATTTT 740

Qy    181 AAATATTGTCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGG 240
            ||||   ::|:  |||       ||:| :|:  :||:|       | :   ||:|
Db    741 TTTTATTTAAWWAWTTAAAATTAWTWATWAAWAWTTWTAWAAAAAAAAATNWTATAAWATT 800

Qy    241 AGACTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTT 287
          :    |||:| | :: |  :  || | :  | | : |  : | |||
Db    801 WWKAATAATWTATAWWTTTAWATATWATRTAAATTWTAATWYTTTTT 847
```

RESULT 12
AZ548198/c
LOCUS AZ548198 853 bp DNA GSS 14-NOV-2000
DEFINITION ENTFW05TF Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
ACCESSION AZ548198
VERSION AZ548198.1 GI:11171567
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 853)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica HM1:IMSS sheared DNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 18
High quality sequence stop: 838.
FEATURES Location/Qualifiers
source 1. .853
/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOS1; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a

tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999)."
BASE COUNT 222 a 127 c 68 g 436 t
ORIGIN

Query Match 6.0%; Score 51.8; DB 245; Length 853;
Best Local Similarity 50.2%; Pred. No. 0.083;
Matches 128; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

```
Qy    444 TAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTATTTAAACATCCAATT 503
          | || || | ||||| | | | | ||||| ||||||| || || |
Db    425 TTAATATATTAAACATAAACAACATTTCAGAATAAAGAATGAATATTTAACAATGAAAGT 366

Qy    504 CATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATCTGCCAGTTGCTTTGCTG 563
            |   | | |||||| | || |||| || | | || |
Db    365 GTTGATTTATTTAATATTTGTTCAACATAAATTCAATAAAACATATATTTTGATAATAT 306

Qy    564 ATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAAAAACCCACACAATGT 623
          || || ||| |||| | | || |||||||||| |||||||  | ||
Db    305 ATAAATGGAATATTAAAGGAATATACAATACTCAAGAACTAATAAAAAAGAAATAAAAAG 246

Qy    624 GAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATCTACATAAGCAGAAGG 683
          | |     | | ||  | || | || ||||| | || | | |  |||
Db    245 AATGAATAAATGAAGATAAATTTGATAAATAATACAAGAAAGAAATATAAACGACAAAGA 186

Qy    684 AAGAGAAATGAAATA 698
          || |||||| | | |
Db    185 AAAAGAAATAATAGA 171
```

RESULT 13
AZ686179/c
LOCUS AZ686179 868 bp DNA GSS 14-DEC-2000
DEFINITION ENTJI41TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
ACCESSION AZ686179
VERSION AZ686179.1 GI:11823325
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 868)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica HM1:IMSS sheared DNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 74
High quality sequence stop: 779.
FEATURES Location/Qualifiers
source 1. .868
/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOS1; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999)."
BASE COUNT 214 a 128 c 82 g 444 t
ORIGIN

Query Match 6.0%; Score 51.8; DB 248; Length 868;
Best Local Similarity 50.2%; Pred. No. 0.083;
Matches 128; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

```
Qy    444 TAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTATTTAAACATCCAATT 503
          | || || | ||||| | | | | ||||| ||||||| || || |
Db    419 TTAATATATTAAACATAAACAACATTTCAGAATAAAGAATGAATATTTAACAATGAAAGT 360

Qy    504 CATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATCTGCCAGTTGCTTTGCTG 563
            |   | | |||||| | | || |||| || | | || |
Db    359 GTTGATTTATTTAATATTTTGTTCAACATAAATTCAATAAAACATATATTTTGATAATAT 300

Qy    564 ATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAAAAACCCACACAATGT 623
          || || |||  |||| | | || |||||||||| |||||||  | ||
Db    299 ATAAATGGAATATTAAAGGAATATACAATACTCAAGAACTAATAAAAAAGAAATAAAAAG 240

Qy    624 GAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATCTACATAAGCAGAAGG 683
          | |     | | ||  | || | || ||||| | || | | |  |||
Db    239 AATGAATAAATGAAGATAAATTTGATAAATAATACAAGAAAGAAATATAAACGACAAAGA 180

Qy    684 AAGAGAAATGAAATA 698
          || |||||| | | |
Db    179 AAAAGAAATAATAGA 165
```

RESULT 14
AZ681140/c
LOCUS AZ681140 881 bp DNA GSS 14-DEC-2000
DEFINITION ENTMC34TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
ACCESSION AZ681140
VERSION AZ681140.1 GI:11818286
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 881)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica HM1:IMSS sheared DNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 35
High quality sequence stop: 747.
FEATURES Location/Qualifiers
source 1. .881
/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"

```
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT      264 a    123 c     59 g    435 t
ORIGIN


  Query Match             6.0%;  Score 51.8;  DB 247;  Length 881;
  Best Local Similarity   50.2%;  Pred. No. 0.083;
  Matches  128;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy    444 TAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTATTTAAACATCCAATT 503
          | || || | |||||  |  |       |   |   |||||  |||||||  ||  || |
Db    619 TTAATATATTAAACATAAACAACATTTCAGAATAAAGAATGAATATTTAACAATGAAAGT 560

Qy    504 CATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATCTGCCAGTTGCTTTGCTG 563
          |    |  | ||||||    |  | ||    ||||  || |     | ||   |
Db    559 GTTGATTTATTTAATATTTTGTTCAACATAAATTCAATAAAACATATATTTTGATAATAT 500

Qy    564 ATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAAAAACCCACACAATGT 623
          ||  || |||   ||||  |  |  ||  |||||||||| |||||||   |   ||
Db    499 ATAAATGGAATATTAAAGGAATATACAATACTCAAGAACTAATAAAAAAGAAATAAAAAG 440

Qy    624 GAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATCTACATAAGCAGAAGG 683
          | |      |  | ||   |   ||  |  ||  |||||  |  || | | |   |||
Db    439 AATGAATAAATGAAGATAAATTTGATAAATAATACAAGAAAGAAATATAAACGACAAAGA 380

Qy    684 AAGAGAAATGAAATA 698
          || |||||| | | |
Db    379 AAAAGAAATAATAGA 365


RESULT  15
AZ542627
LOCUS       AZ542627      899 bp    DNA             GSS       14-NOV-2000
DEFINITION  ENTCX21TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   AZ542627
VERSION     AZ542627.1  GI:11160786
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 899)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 93
            High quality sequence stop: 858.
FEATURES             Location/Qualifiers
     source          1. .899
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT      429 a     61 c    150 g    259 t
ORIGIN


  Query Match             6.0%;  Score 51.8;  DB 245;  Length 899;
  Best Local Similarity   50.2%;  Pred. No. 0.083;
  Matches  128;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy    444 TAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTATTTAAACATCCAATT 503
          | || || | |||||  |  |       |   |   |||||  |||||||  ||  || |
Db    371 TTAATATATTAAACATAAACAACATTTCAGAATAAAGAATGAATATTTAACAATGAAAGT 430

Qy    504 CATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATCTGCCAGTTGCTTTGCTG 563
          |    |  | ||||||    |  | ||    ||||  || |     | ||   |
Db    431 GTTGATTTATTTAATATTTTGTTCAACATAAATTCAATAAAACATATATTTTGATAATAT 490

Qy    564 ATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAAAAACCCACACAATGT 623
          ||  || |||   ||||  |  |  ||  |||||||||| |||||||   |   ||
Db    491 ATAAATGGAATATTAAAGGAATATACAATACTCAAGAACTAATAAAAAAGAAATAAAAAG 550

Qy    624 GAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATCTACATAAGCAGAAGG 683
          | |      |  | ||   |   ||  |  ||  |||||  |  || | | |   |||
Db    551 AATGAATAAATGAAGATAAATTTGATAAATAATACAAGAAAGAAATATAAACGACAAAGA 610

Qy    684 AAGAGAAATGAAATA 698
          || |||||| | | |
Db    611 AAAAGAAATAATAGA 625


Search completed: November 20, 2001, 16:21:54
Job time: 4448 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2001, 14:27:41 ; Search time 1472.19 Seconds
(without alignments)
9046.205 Million cell updates/sec

Title: US-09-308-080-1
Perfect score: 861
Sequence: 1 TGTTAATGAAGATAAATATT.....AGTGGGAAATAATTTATTAA 861

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	775.2	90.0	180286	80	AL356574	AL356574 Homo sapi
	2	608.6	70.7	626	97	HSU57655	U57655 Human dihyd
	3	488.2	56.7	854	93	HSDYPDGEN	X95670 H.sapiens D
	4	171.2	19.9	3951	97	HSU09178	U09178 Human dihyd
	5	171.2	19.9	3957	9	AR027740	AR027740 Sequence
	6	169.6	19.7	3078	85	AB003063	AB003063 Homo sapi
	7	169.6	19.7	4409	97	HSU20938	U20938 Human lymph
	8	153.4	17.8	4421	7	BTU20981	U20981 Bos taurus

	9	150.2	17.4	4441	7	SSU09179 U09179 Sus scrofa
	10	150.2	17.4	4447	9	AR027741 AR027741 Sequence
	11	145.2	16.9	4358	94	D85035 D85035 Rattus norv
	12	70.6	8.2	43803	6	CELC25F6 U39742 Caenorhabdi
	13	70	8.1	24429	63	AC014157 AC014157 Drosophil
c	14	70	8.1	303092	4	AE003446 AE003446 Drosophil
	15	68.4	7.9	2348	6	DMU65491 U65491 Drosophila
	16	55.4	6.4	117674	84	DMBR30C13 AL122025 Drosophil
c	17	51	5.9	141953	71	AC044793 AC044793 Homo sapi
c	18	50	5.8	156060	60	AC004153 AC004153 Plasmodiu
	19	49.8	5.8	150803	61	AC010228 AC010228 Homo sapi
	20	49.6	5.8	1017	6	DDCPRO2U X04775 Dictyosteli
c	21	49.4	5.7	147490	90	AL445207 AL445207 Human DNA
	22	49.4	5.7	160728	70	AC026816 AC026816 Homo sapi
c	23	49.4	5.7	161025	66	AC021276 AC021276 Homo sapi
	24	49.4	5.7	164450	82	AP000794 AP000794 Homo sapi
	25	49.2	5.7	168674	69	AC025645 AC025645 Homo sapi
c	26	48.8	5.7	94536	15	T14P8 AF069298 Arabidops
	27	48.8	5.7	194892	13	ATCHRIV6 AL161494 Arabidops
c	28	48.8	5.7	217327	79	AL162723 AL162723 Homo sapi
c	29	48.2	5.6	1149	5	AF044859 AF044859 Eurema da
c	30	48.2	5.6	12029	4	AE001412 AE001412 Plasmodiu
	31	48	5.6	590	96	PARMTDI4D K00907 paramecium
	32	48	5.6	692	96	PARMTDI4H K00911 paramecium
	33	48	5.6	726	96	PARMTDI4C K00906 paramecium
	34	48	5.6	760	96	PARMTDI4A K00904 paramecium
	35	48	5.6	40469	6	MIPAGEN X15917 Paramecium
	36	48	5.6	47573	5	AF030694 AF030694 Plasmodiu
c	37	47.4	5.5	112519	85	AB041992 AB041992 Homo sapi
	38	47.4	5.5	148965	72	AC060784 AC060784 Homo sapi
	39	47.4	5.5	165823	71	AC027625 AC027625 Homo sapi
c	40	47.4	5.5	177920	64	AC016700 AC016700 Homo sapi
c	41	47.4	5.5	300000	91	AP002529 AP002529 Homo sapi
c	42	47.2	5.5	86827	96	PFMAL3P5 AL034556 Plasmodiu
c	43	47	5.5	7218	10	I66494 I66494 Sequence 14
	44	47	5.5	65691	96	PFMAL3P1 Z97348 Plasmodium
	45	46.8	5.4	87509	12	AC004482 AC004482 Arabidops

ALIGNMENTS

RESULT 1
AL356574/c
LOCUS AL356574 180286 bp DNA HTG 23-JAN-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-359C24, *** SEQUENCING IN PROGRESS ***, 26 unordered pieces.
ACCESSION AL356574
VERSION AL356574.3 GI:9797568
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 180286)
AUTHORS Plumb,B.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Aug 12, 2000 this sequence version replaced gi:9213933.
---------- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
---------- Project Information
Center project name: bA359C24
---------- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 166105 bases at least Q40
Consensus quality: 171672 bases at least Q30
Consensus quality: 174563 bases at least Q20
Insert size: 177786; sum-of-contigs
Insert size: 186027; 5.1% error; agarose-fp
Quality coverage: 3.18x in Q20 bases; sum-of-contigs Quality coverage: 3.10x in Q20 bases; agarose-fp
----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 15090: contig of 15090 bp in length
* 15091 15190: gap of 100 bp
* 15191 23545: contig of 8355 bp in length
* 23546 23645: gap of 100 bp
* 23646 26766: contig of 3121 bp in length
* 26767 26866: gap of 100 bp
* 26867 38453: contig of 11587 bp in length
* 38454 38553: gap of 100 bp
* 38554 45878: contig of 7325 bp in length
* 45879 45978: gap of 100 bp
* 45979 51984: contig of 6006 bp in length
* 51985 52084: gap of 100 bp
* 52085 58204: contig of 6120 bp in length
* 58205 58304: gap of 100 bp
* 58305 71630: contig of 13326 bp in length
* 71631 71730: gap of 100 bp
* 71731 75338: contig of 3608 bp in length
* 75339 75438: gap of 100 bp
* 75439 85213: contig of 9775 bp in length
* 85214 85313: gap of 100 bp
* 85314 88517: contig of 3204 bp in length
* 88518 88617: gap of 100 bp
* 88618 92454: contig of 3837 bp in length
* 92455 92554: gap of 100 bp
* 92555 98200: contig of 5646 bp in length
* 98201 98300: gap of 100 bp
* 98301 111029: contig of 12729 bp in length
* 111030 111129: gap of 100 bp
* 111130 117311: contig of 6182 bp in length
* 117312 117411: gap of 100 bp
* 117412 121093: contig of 3682 bp in length
* 121094 121193: gap of 100 bp
* 121194 137336: contig of 16143 bp in length
* 137337 137436: gap of 100 bp
* 137437 141486: contig of 4050 bp in length
* 141487 141586: gap of 100 bp
* 141587 146703: contig of 5117 bp in length
* 146704 146803: gap of 100 bp
* 146804 149034: contig of 2231 bp in length
* 149035 149134: gap of 100 bp
* 149135 154764: contig of 5630 bp in length
* 154765 154864: gap of 100 bp
* 154865 159990: contig of 5126 bp in length
* 159991 160090: gap of 100 bp
* 160091 163742: contig of 3652 bp in length
* 163743 163842: gap of 100 bp
* 163843 166322: contig of 2480 bp in length
* 166323 166422: gap of 100 bp
* 166423 176700: contig of 10278 bp in length
* 176701 176800: gap of 100 bp
* 176801 180286: contig of 3486 bp in length.
FEATURES Location/Qualifiers
source 1. .180286
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-359C24"
/clone_lib="RPCI-11.2"

```
     misc_feature    1. .15090
                     /note="assembly_fragment:00687
                     clone_end:SP6
                     vector_side:left"
     misc_feature    15191. .23545
                     /note="assembly_fragment:00038
                     fragment_chain:1"
     misc_feature    23646. .26766
                     /note="assembly_fragment:00972
                     fragment_chain:1"
     misc_feature    26867. .38453
                     /note="assembly_fragment:01030
                     fragment_chain:1"
     misc_feature    38554. .45878
                     /note="assembly_fragment:00178
                     fragment_chain:2"
     misc_feature    45979. .51984
                     /note="assembly_fragment:00498
                     fragment_chain:2"
     misc_feature    52085. .58204
                     /note="assembly_fragment:00217
                     fragment_chain:3"
     misc_feature    58305. .71630
                     /note="assembly_fragment:01029
                     fragment_chain:3"
     misc_feature    71731. .75338
                     /note="assembly_fragment:00265
                     fragment_chain:4"
     misc_feature    75439. .85213
                     /note="assembly_fragment:01189
                     fragment_chain:4"
     misc_feature    85314. .88517
                     /note="assembly_fragment:00320
                     fragment_chain:5"
     misc_feature    88618. .92454
                     /note="assembly_fragment:00693
                     fragment_chain:5"
     misc_feature    92555. .98200
                     /note="assembly_fragment:00834
                     fragment_chain:6"
     misc_feature    98301. .111029
                     /note="assembly_fragment:00385
                     fragment_chain:6"
     misc_feature    111130. .117311
                     /note="assembly_fragment:00932
                     fragment_chain:7"
     misc_feature    117412. .121093
                     /note="assembly_fragment:01129
                     fragment_chain:7"
     misc_feature    121194. .137336
                     /note="assembly_fragment:00160"
     misc_feature    137437. .141486
                     /note="assembly_fragment:00194"
     misc_feature    141587. .146703
                     /note="assembly_fragment:00234"
     misc_feature    146804. .149034
                     /note="assembly_fragment:00529"
     misc_feature    149135. .154764
                     /note="assembly_fragment:00590"
     misc_feature    154865. .159990
                     /note="assembly_fragment:01001"
     misc_feature    160091. .163742
                     /note="assembly_fragment:01119"
     misc_feature    163843. .166322
                     /note="assembly_fragment:01294"
     misc_feature    166423. .176700
                     /note="assembly_fragment:01323"
     misc_feature    176801. .180286
                     /note="assembly_fragment:01326"
BASE COUNT    57303 a  32087 c  31562 g  56812 t   2522 others
ORIGIN
```

```
  Query Match             90.0%;  Score 775.2;  DB 80;  Length 180286;
  Best Local Similarity   95.8%;  Pred. No. 2.1e-142;
  Matches  828;  Conservative    0;  Mismatches   33;  Indels    3;  Gaps    3;

Qy      1 TGTTAATGAAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGT 60
          |||||||||| |||||||||||| ||||| |||| |||||||||||  ||||||||||
Db  50426 TGTTAATGAATATAAATATTTTTATTTTTGCGCTATTCTAAACCTAGAATTACAAGAAGT 50367

Qy     61 AATTTATCTGGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTT 120
          |||||| ||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db  50366 AATTTAGCTAAAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTT 50307

Qy    121 CAATTCTAATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATA 180
          |||||||||||||||||||||||||||||||||||||||| ||||||| ||||||||||
Db  50306 CAATTCTAATTTAATGTATATTAAAAATTCCTCTGCAAAAATGTGAGAAGGGACCTCATA 50247

Qy    181 AAATATTGTCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  50246 AAATATTGTCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGG 50187

Qy    241 AGACTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCAT 300
          |||||||||||||||||||||||| |||||||||||||||||||||||| ||||||
Db  50186 AGACTCAATATCTTTACTCTTTCATCAGGACATTGTGACAAATGTTTCCCCCAGAATCAT 50127

Qy    301 CCGGGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  50126 CCGGGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGA 50067

Qy    361 GCTCATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  50066 GCTCATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGA 50007

Qy    421 CTTCCCAGACAACGTAAGTGTGA-TAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGT 479
          ||| |||||||||||||||||| | || ||||||||||||||||||||||||||||||
Db  50006 CTTTCCAGACAACGTAAGTGTGATTTAACATCTAAAACAAGAGAATTGGCATAAGTTGGT 49947

Qy    480 GAATGTTTATTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCA 539
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db  49946 GAATGTTTATTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATTA 49887

Qy    540 ACGAATCTGCCAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAG 599
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  49886 AAGAATCTGCCAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAG 49827

Qy    600 AACTCATAAAAACCCACACAATGTGAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGA 659
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db  49826 AACTCATAAAAACCCACACAATGTGAAGCTTTGTTATAAATGGGTGCCATGTAAGATGGA 49767

Qy    660 AGAAGTATCTACATAAGCAGAAGGAAGAGAAATGAAATACTCATTTTATTGAGTTGGCCC 719
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  49766 AGAAGTATCTACATAAGCAGAAGGAAGAGAAATGAAATACTCATTTTATTGAGTTGGTTT 49707

Qy    720 CCACTGTATGTGGCTGGTATTTATGAAGGTGATGACC-CAGGAAGAAATTGTAAACTATA 778
          |||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db  49706 TCACTGTATGTGGCTGGTATTTATGAAGGTGATGACCTTAGGAAGAAATTGTAAACTATA 49647

Qy    779 AACCACTCCAAATATAAACCCGAGGCAGAAGCAGCATATC-TCCTATGAAGCCTGTATTT 837
          || || |  |||||||||| |||||||||||||||||| |||||||||| || |||||
Db  49646 AATCATTTTAAATATAAACCTGAGGCAGAAGCAGCATATCTTCCTATGAAGTCTATATTT 49587

Qy    838 ACTCAGTGGGAAATAATTTATTAA 861
            ||||||||||||||||||||||
Db  49586 TTTCAGTGGGAAATAATTTATTAA 49563
```

```
RESULT   2
HSU57655
LOCUS       HSU57655      626 bp    DNA             PRI       01-SEP-1996
DEFINITION  Human dihydropyrimidine dehydrogenase (DPYD) gene, partial cds.
ACCESSION   U57655
VERSION     U57655.1  GI:1518606
KEYWORDS    .
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 626)
  AUTHORS   Fernandez-Salguero,P.M., Wei,X., Jones,S., Idle,J.R. and
            Gonzalez,F.J.
  TITLE     Lack of dihydropyrimidine dehydrogenase activity and
            thymine-uraciluria caused by a common splice mutation
  JOURNAL   Hum. Mutat. (1996) In press
REFERENCE   2  (bases 1 to 626)
  AUTHORS   Fernandez-Salguero,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-1996) Pedro M. Fernandez-Salguero, National
            Institutes of Health, 9000 Rockville Pike, Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .626
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            179. .343
                     /gene="DPYD"
     CDS             <179. .>343
                     /gene="DPYD"
                     /function="uracil and thymine metabolism"
                     /codon_start=1
                     /evidence=experimental
                     /product="dihydropyrimidine dehydrogenase"
                     /protein_id="AAB07049.1"
                     /db_xref="GI:1518607"
                     /translation="DIVTNVSPRIIRGTTSGPMYGPGQSSFLNIELISEKTAAYWCQS
                     VTELKADFPDN"
BASE COUNT      221 a    100 c    117 g    188 t
ORIGIN
```

Query Match 70.7%; Score 608.6; DB 97; Length 626;
Best Local Similarity 99.2%; Pred. No. 9.2e-110;
Matches 622; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

```
Qy      90 TTTACCTTTTTATTTGCAAGTAGTTTATGTTCAATTCTAATTTAATGTATATTAAAAATT 149
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TTTACCTTTTTATTTGCAAGTAGTTTATGTTCAATTCTAATTTAATGTATATTAAAAATT 60

Qy     150 CCTCTGCAAATATGTGAGGAGGGACCTCATAAAATATTGTCATATGGAAATGAGCAGATA 209
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db      61 CCTCTGCAAATATGTGAGGAGGGACCTCAT-AAATATTGTCATATGGAAATGAGCAGATA 119

Qy     210 ATAAAGATTATAGCTTTTCTTTGTCAAAAGGAGACTCAATATCTTTACTCTTTCATGAGG 269
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     120 ATAAAGATTATAGCTTTTCTTTGTCAAAAGGAGACTCAATATCTTTACTCTTTCATCAGG 179

Qy     270 ACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAACCACCTCTGGCCCCATGTATG 329
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     180 ACATTGTGACAAATGTTTCCCCCAGAATCATCCGGGGAACCACCTCTGGCCCCATGTATG 239

Qy     330 GCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCAGTGAGAAAACGGCTGCATATT 389
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     240 GCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCAGTGAGAAAACGGCTGCATATT 299

Qy     390 GGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAGACAACGTAAGTGTGATAAAAA 449
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     300 GGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAGACAACGTAAGTGTGATAAAAA 359

Qy     450 TCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTATTTAAACATCCAATTCATAGG 509
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     360 TCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTATTTAAACATCCAATTCATAGG 419

Qy     510 CTTATAAATATTAATGTGTATATTTTATCAACGAATCTGCCAGTTGCTTTGCTGATGCAT 569
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db     420 CTTATAAATATTAATGTGTATATTTTATCAAAGAATCTGCCAGTTGCTTTGCTGATGCAT 479

Qy     570 AGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAAAAACCCACACAATGTGAAGCT 629
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     480 AGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAAAAACCCACACAATGTGAAGCT 539

Qy     630 CTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATCTACATAAGCAGAAGGAAGAGA 689
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     540 CTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATCTACATAAGTAGAAGGAAGAGA 599

Qy     690 AATGAAATACTCATTTTATTGAGTTGG 716
           |||||||||||||||||||||||||||
Db     600 AATGAAATACTCATTTTATTGAGTTGG 626
```

```
RESULT   3
HSDYPDGEN
LOCUS       HSDYPDGEN     854 bp    DNA             PRI       17-FEB-1997
DEFINITION  H.sapiens DYPD gene, partial sequence.
ACCESSION   X95670
VERSION     X95670.1  GI:1246752
KEYWORDS    dihydropyrimidine dehydrogenase; DYPD gene; exon X; mutation.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 854)
  AUTHORS   Vreken,P., Van Kuilenburg,A.B., Meinsma,R., Smit,G.P., Bakker,H.D.,
            De Abreu,R.A. and van Gennip,A.H.
  TITLE     A point mutation in an invariant splice donor site leads to exon
            skipping in two unrelated Dutch patients with dihydropyrimidine
            dehydrogenase deficiency
  JOURNAL   J. Inherit. Metab. Dis. 19 (5), 645-654 (1996)
  MEDLINE   97047101
REFERENCE   2  (bases 1 to 854)
  AUTHORS   Vreken,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-1996) P. Vreken, Academic Med.Center, Univ.of
            Amsterdam, Dep of Pediatrics and Clinical Chemistry, F0-224, P.O.
            Box 22700, NL-1100 DE Amsterdam, NETHERLANDS
COMMENT     Related sequences:-
            Meinsma et al, DNA Cell Biol. 14:1-6 (1995), U20938, and Yokota H.
            et al, J. Biol. Chem. 269:23192-23196 (1994).
FEATURES             Location/Qualifiers
     source          1. .854
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /tissue_type="liver"
                     /map="p22"
     intron          <1. .568
     gene            569. .734
                     /gene="DPYD"
     CDS             <569. .>734
                     /gene="DPYD"
                     /EC_number="1.3.1.2"
                     /note="exon X"
                     /codon_start=2
                     /product="dihydropyrimidine dehydrogenase (NADP+)"
                     /protein_id="CAA64973.1"
                     /db_xref="GI:1246753"
                     /db_xref="SWISS-PROT:Q12882"
                     /translation="DIVTNVSPRIIRGTTSGPMYGPGQSSFLNIELISEKTAAYWCQS
                     VTELKADFPDN"
     exon            569. .734
                     /gene="DPYD"
     variation       725
                     /gene="DPYD"
                     /replace="c"
     intron          735. .>854
     variation       735
                     /note="changes the splice donor sequence and leads to exon
                     skipping"
                     /replace="a"
BASE COUNT      301 a    126 c    130 g    297 t
ORIGIN
```

```
  Query Match            56.7%;  Score 488.2;  DB 93;  Length 854;
  Best Local Similarity  96.2%;  Pred. No. 3.5e-86;
  Matches  532;  Conservative    0;  Mismatches   18;  Indels    3;  Gaps    3;

Qy      1 TGTTAATGAAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGT 60
          |||||||||| |||||||||||| ||||| |||| |||||||||||  ||||||||||
Db    304 TGTTAATGAATATAAATATTTTTATTTTTGCGCTATTCTAAACCTAGAATTACAAGAAGT 363

Qy     61 AATTTATCTGGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTT 120
          |||||| ||  |||||||||||||||||||||| ||||||||||||||||||||||||
Db    364 AATTTAGCTAAAGCTAACAAATACTTTATTTTACC-TTTTATTTGCAAGTAGTTTATGTT 422

Qy    121 CAATTCTAATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATA 180
          ||||||||||||||||||||||||||||||||||| ||||||| ||||||||||
Db    423 CAATTCTAATTTAATGTATATTAAAAATTCCTCTGCAAAAATGTGAGAAGGGACCTCATA 482

Qy    181 AAATATTGTCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGG 240
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    483 AAATA-TGTCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGG 541

Qy    241 AGACTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCAT 300
          |||||||||||||||||||||||| |||||||||||||||||||||| ||||||
Db    542 AGACTCAATATCTTTACTCTTTCATCAGGACATTGTGACAAATGTTTCCCCCAGAATCAT 601

Qy    301 CCGGGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    602 CCGGGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGA 661

Qy    361 GCTCATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    662 GCTCATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGA 721

Qy    421 CTTCCCAGACAACGTAAGTGTGA-TAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGT 479
          ||| |||||||||||||||||| | || ||||||||||||||||||||||||||||||
Db    722 CTTTCCAGACAACGTAAGTGTGATTTAACATCTAAAACAAGAGAATTGGCATAAGTTGGT 781

Qy    480 GAATGTTTATTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCA 539
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    782 GAATGTTTATTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATTA 841

Qy    540 ACGAATCTGCCAG 552
          | ||||||||||||
Db    842 AAGAATCTGCCAG 854


RESULT  4
HSU09178
LOCUS       HSU09178     3951 bp    mRNA            PRI       28-DEC-1994
DEFINITION  Human dihydropyrimidine dehydrogenase mRNA, complete cds.
ACCESSION   U09178
VERSION     U09178.1  GI:558304
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3951)
  AUTHORS   Yokota,H., Fernandez-Salguero,P., Furuya,H., Lin,K., McBride,O.W.,
            Podschun,B., Schnackerz,K.D. and Gonzalez,F.J.
  TITLE     cDNA cloning and chromosome mapping of human dihydropyrimidine
            dehydrogenase, an enzyme associated with 5-fluorouracil toxicity
            and congenital thymine uraciluria
  JOURNAL   J. Biol. Chem. 269 (37), 23192-23196 (1994)
  MEDLINE   94365020
REFERENCE   2  (sites)
  AUTHORS   Eggink,G., Engel,H., Vriend,G., Terpstra,P. and Witholt,B.
  TITLE     Rubredoxin reductase of Pseudomonas oleovorans. Structural
            relationship to other flavoprotein oxidoreductases based on one NAD
            and two FAD fingerprints
  JOURNAL   J. Mol. Biol. 212 (1), 135-142 (1990)
  MEDLINE   90204534
REFERENCE   3  (sites)
  AUTHORS   Porter,D.J., Chestnut,W.G., Merrill,B.M. and Spector,T.
  TITLE     Mechanism-based inactivation of dihydropyrimidine dehydrogenase by
            5-ethynyluracil
  JOURNAL   J. Biol. Chem. 267 (8), 5236-5242 (1992)
  MEDLINE   92184771
REFERENCE   4  (sites)
  AUTHORS   Dupuis,A., Skehel,J.M. and Walker,J.E.
  TITLE     A homologue of a nuclear-coded iron-sulfur protein subunit of
            bovine mitochondrial complex I is encoded in chloroplast genomes
  JOURNAL   Biochemistry 30 (11), 2954-2960 (1991)
  MEDLINE   91175743
REFERENCE   5  (sites)
  AUTHORS   Wierenga,R.K., De Maeyer,M.C.H. and Hol,W.G.J.
  TITLE     Interaction of pyrophosphatase moieties with alfa-helixes in
            dinucleotide binding proteins
  JOURNAL   Biochemistry 24, 1346-1357 (1985)
REFERENCE   6  (bases 1 to 3951)
  AUTHORS   Gonzalez,F.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-APR-1994) Frank J. Gonzalez, National Cancer
            Institute, National Institutes of Health, 9000 Rockville Pike,
            Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .3951
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="liver"
     CDS             82. .3159
                     /codon_start=1
                     /product="dihydropyrimidine dehydrogenase"
                     /protein_id="AAA57474.1"
                     /db_xref="GI:558305"
                     /translation="MAPVLSKDSADIESILALNPRTQTHATLCSTSAKKLDKKHWKRN
                     PDKNCFNCEKLENNFDDIKHTTLGERGALREAMRCLKCADAPCQKSCPTNLDIKSFIT
                     SIANKNYYGAAKMIFSDNPLGLTCGMVCPTSDLCVGGCNLYATEEGPINIGGLQQFAT
                     EVFKAMSIPQIRNPSLPPPEKMSEAYSAKIALFGAGPASISCASFLARLGYSDITIFE
                     KQEYVGGLSTSEIPQFRLPYDVVNFEIELMKDLGVKIICGKSLSVNEMTLSTLKEKGY
                     KAAFIGIGLPEPNKDAIFQGLTQDQGFYTSKDFLPLVAKGSKAGMCACHSPLPSIRGV
                     VIVLGAGDTAFDCATSALRCGARRVFIVFRKGFVNIRAVPEEMELAKEEKCEFLPFLS
                     PRKVIVKGGRIVAMQFVRTEQDETGKWNEDEDQMVHLKADVVISAFGSVLSDPKVKEA
                     LSPIKFNRWGLPEVDPETMQTSEAWVFAGGDVVGLANTTVESVNDGKQASWYIHKYVQ
                     SQYGASVSAKPELPLFYTPIDLVDISVEMAGLKFINPFGLASATPATSTSMIRRAFEA
                     GWGFALTKTFSLDKDIVTNVSPRIIRGTTSGPMYGPGQSSFLNIELISEKTAAYWCQS
                     VTELKADFPDNIVIASIMCSYNKNDWTELAKKSEDSGADALELNLSCPHGMGERGMGL
                     ACGQDPELVRNICRWVRQAVQIPFFAKLTPNVTDIVSIARAAKEGGANGVTATNTVSG
                     LMGLKSDGTPWPAVGIAKRTTYGGVSGTAIRPIALRAVTSIARALPGFPILATGGIDS
                     AESGLQFLHSGASVLQVCSAIQNQDFTVIEDYCTGLKALLYLKSIEELQDWDGQSPAT
                     VSHQKGKPVPRIAELMDKKLPSFGPYLEQRKKIIAENKIRLKEQNVAFSPLKRSCFIP
                     KRPIPTIKDVIGKALQYLGTFGELSNVEQVVAMIDEEMCINCGKCYMTCNDSGYQAIQ
                     FDPETHLPTITDTCTGCTLCLSVCPIVDCIKMVSRTTPYEPKRGVPLSVNPVC"
     misc_feature    1084. .1134
                     /citation=[5]
                     /function="catalytic cofactor NADPH/NADP binding site"
     misc_feature    1492. .1524
                     /citation=[2]
                     /function="electron transfer center, FAD binding site"
     misc_feature    2062. .2175
                     /citation=[3]
                     /function="uracil (substrate) binding site"
     misc_feature    2938. .2973
                     /standard_name="iron-sulfur center"
                     /citation=[4]
                     /function="catalytic cofactor [4Fe-4S] binding site"
BASE COUNT     1153 a    785 c    896 g   1117 t
ORIGIN


  Query Match            19.9%;  Score 171.2;  DB 97;  Length 3951;
  Best Local Similarity  93.2%;  Pred. No. 4.2e-24;
  Matches  179;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy    244 CTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCG 303
          | | | | ||||     || || |||||||||||||||||||||||||| |||||||||
Db   1797 CACCAAAACTTTCTCTCTTGATAAGGACATTGTGACAAATGTTTCCCCCAGAATCATCCG 1856
```

```
Qy     304 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1857 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 1916

Qy     364 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 423
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1917 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 1976

Qy     424 CCCAGACAACGT 435
           |||||||||| |
Db    1977 CCCAGACAACAT 1988
```

```
RESULT   5
AR027740
LOCUS       AR027740     3957 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 1 from patent US 5856454.
ACCESSION   AR027740
VERSION     AR027740.1  GI:5938560
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3957)
  AUTHORS   Gonzalez,F.J. and Fernandez-Salguero,P.
  TITLE     CDNA for human and pig dihydropyrimidine dehydrogenase
  JOURNAL   Patent: US 5856454-A 1 05-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .3957
                     /organism="unknown"
BASE COUNT     1156 a    787 c    897 g   1117 t
ORIGIN
```

```
  Query Match           19.9%;  Score 171.2;  DB 9;  Length 3957;
  Best Local Similarity   93.2%;  Pred. No. 4.2e-24;
  Matches  179;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy     244 CTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCG 303
           | | | | ||||     || || ||||||||||||||||||||||||| |||||||||
Db    1803 CACCAAAACTTTCTCTCTTGATAAGGACATTGTGACAAATGTTTCCCCCAGAATCATCCG 1862

Qy     304 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1863 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 1922

Qy     364 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 423
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1923 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 1982

Qy     424 CCCAGACAACGT 435
           |||||||||| |
Db    1983 CCCAGACAACAT 1994
```

```
RESULT   6
AB003063
LOCUS       AB003063     3078 bp    mRNA            PRI       21-JAN-2000
DEFINITION  Homo sapiens mRNA for dihydropyrimidine dehydrogenase, complete
            cds.
ACCESSION   AB003063
VERSION     AB003063.1  GI:6729337
KEYWORDS    dihydropyrimidine dehydrogenase.
SOURCE      Homo sapiens male lymphocyte cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ogura,K., Nishiyama,T., Takubo,H., Kato,A., Okuda,H., Arakawa,K.,
            Fukushima,M., Nagayama,S., Kawaguchi,Y. and Watabe,T.
  TITLE     Suicidal inactivation of human dihydropyrimidine dehydrogenase by
            (E)-5-(2-bromovinyl)uracil derived from the antiviral, sorivudine
  JOURNAL   Cancer Lett. 122 (1-2), 107-113 (1998)
  MEDLINE   98124145
  REMARK    Erratum:[[published erratum appears in Cancer Lett 1998 Jun
            19;128(2):229]]
REFERENCE   2  (bases 1 to 3078)
  AUTHORS   Ogura,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-1997) to the DDBJ/EMBL/GenBank databases.
            Kenichiro Ogura, Tokyo University of Pharmacy and Life Science,
            Department of Drug Metabolism and Molecular Toxicology; 1432-1
            Horinouchi, Hachioji, Tokyo 192-0392, Japan
            (E-mail:ogurak@ps.toyaku.ac.jp, Tel:+81-426-76-4518,
            Fax:+81-426-76-4517)
FEATURES             Location/Qualifiers
     source          1. .3078
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="lymphocyte"
                     /sex="male"
     CDS             1. .3078
                     /codon_start=1
                     /product="dihydropyrimidine dehydrogenase"
                     /protein_id="BAA89789.1"
                     /db_xref="GI:6729338"
                     /translation="MAPVLSKDSADIESILALNPRTQTHATLCSTSAKKLDKKHWKRN
                     PDKNCFNCEKLENNFDDIKHTTLGERGALREAMRCLKCADAPCQKSCPTNLDIKSFIT
                     SIANKNYYGAAKMIFSDNPLGLTCGMVCPTSDLCVGGCNLYATEEGPINIGGLQQFAT
                     EVFKAMSIPQIRNPSLPPPEKMSEAYSAKIALFGAGPASISCASFLARLGYSDITIFE
                     KQEYVGGLSTSEIPQFRLPYDVVNFEIELMKDLGVKIICGKSLSVNEMTLSTLKEKGY
                     KAAFIGIGLPEPNKDAIFQGLTQDQGFYTSKDFLPLVAKGSKAGMCACHSPLPSIRGV
                     VIVLGAGDTAFDCATSALRCGARRVFIVFRKGFVNIRAVPEEMELAKEEKCEFLPFLS
                     PRKVIVKGGRIVAMQFVRTEQDETGKWNEDEDQMVHLKADVVISAFGSVLSDPKVKEA
                     LSPIKFNRWGLPEVDPETMQTSEAWVFAGGDVVGLANTTVESVNDGKQASWYIHKYVQ
                     SQYGASVSAKPELPLFYTPIDLVDISVEMAGLKFINPFGLASATPATSTSMIRRAFEA
                     GWGFALTKTFSLDKDIVTNVSPRIIRGTTSGPMYGPGQSSFLNIELISEKTAAYWCQS
                     VTELKADFPDNIVIASIMCSYNKNDWTELAKKSEDSGADALELNLSCPHGMGERGMGL
                     ACGQDPELVRNICRWVRQAVQIPFFAKLTPNVTDIVSIARAAKEGGANGVTATNTVSG
                     LMGLKSDGTPWPAVGIAKRTTYGGVSGTAIRPIALRAVTSIARALPGFPILATGGIDS
                     AESGLQFLHSGASVLQVCSAIQNQDFTVIEDYCTGLKALLYLKSIEELQDWDGQSPAT
                     VSHQKGKPVPRIAELMDKKLPSFGPYLEQRKKIIAENKIRLKEQNVAFSPLKRNCFIP
                     KRPIPTIKDVIGKALQYLGTFGELSNVEQVVAMIDEEMCINCGKCYMTCNDSGYQAIQ
                     FDPETHLPTITDTCTGCTLCLSVCPIVDCIKMVSRTTPYEPKRGVPLSVNPVC"
BASE COUNT      875 a    643 c    734 g    826 t
ORIGIN
```

```
  Query Match           19.7%;  Score 169.6;  DB 85;  Length 3078;
  Best Local Similarity   92.7%;  Pred. No. 8.6e-24;
  Matches  178;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy     244 CTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCG 303
           | | | | ||||     || || ||||||||||||||||||||||||| |||||||||
Db    1716 CACCAAAACTTTCTCTCTTGATAAGGACATTGTGACAAATGTTTCCCCCAGAATCATCCG 1775

Qy     304 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1776 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 1835

Qy     364 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 423
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1836 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 1895

Qy     424 CCCAGACAACGT 435
           |||||||||| |
Db    1896 TCCAGACAACAT 1907
```

```
RESULT   7
HSU20938
LOCUS       HSU20938     4409 bp    mRNA            PRI       04-APR-1997
DEFINITION  Human lymphocyte dihydropyrimidine dehydrogenase mRNA, complete
            cds.
ACCESSION   U20938
```

```
VERSION     U20938.1  GI:1926407
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4409)
  AUTHORS   Diasio,R.B., Beavers,T.L. and Carpenter,J.T.
  TITLE     Familial deficiency of dihydropyrimidine dehydrogenase. Biochemical
            basis for familial pyrimidinemia and severe 5-fluorouracil-induced
            toxicity
  JOURNAL   J. Clin. Invest. 81 (1), 47-51 (1988)
  MEDLINE   88087908
REFERENCE   2  (bases 1 to 4409)
  AUTHORS   Lu,Z.H., Zhang,R. and Diasio,R.B.
  TITLE     Purification and characterization of dihydropyrimidine
            dehydrogenase from human liver
  JOURNAL   J. Biol. Chem. 267 (24), 17102-17109 (1992)
  MEDLINE   92381021
REFERENCE   3  (bases 1 to 4409)
  AUTHORS   Johnson,M.R., Albin,N., Shahinian,H. and Diasio,R.B.
  TITLE     Identification of a frameshift in the DNA coding for
            dihydropyrimidine dehydrogenase (DPD) in a DPD deficient patient
            exhibiting 5-fluorouracil toxicity
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 4409)
  AUTHORS   Johnson,M.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-1995) Pharmacology, University of Alabama at
            Birmingham, 1670 University Blvd., Birmingham, AL 35294-0019, USA
COMMENT     On Apr 4, 1997 this sequence version replaced gi:693911.
FEATURES             Location/Qualifiers
     source          1. .4409
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="lymphocyte"
     CDS             102. .3179
                     /EC_number="1.2.1.2"
                     /note="DPD; dihydrouracil dehydrogenase"
                     /codon_start=1
                     /product="dihydropyrimidine dehydrogenase"
                     /protein_id="AAB51366.1"
                     /db_xref="GI:693912"
                     /translation="MAPVLSKDSADIESILALNPRTQTHATLCSTSAKKLDKKHWKRN
                     PDKNCFNCEKLENNFDDIKHTTLGERGALREAMRCLKCADAPCQKSCPTNLDIKSFIT
                     SIANKNYYGAAKMIFSDNPLGLTCGMVCPTSDLCVGGCNLYATEEGPINIGGLQQFAT
                     EVFKAMSIPQIRNPSLPPPEKMSEAYSAKIALFGAGPASISCASFLARLGYSDITIFE
                     KQEYVGGLSTSEIPQFRLPYDVVNFEIELMKDLGVKIICGKSLSVNEMTLSTLKEKGY
                     KAAFIGIGLPEPNKDAIFQGLTQDQGFYTSKDFLPLVAKGSKAGMCACHSPLPSIRGV
                     VIVLGAGDTAFDCATSALRCGARRVFIVFRKGFVNIRAVPEEMELAKEEKCEFLPFLS
                     PRKVIVKGGRIVAMQFVRTEQDETGKWNEDEDQMVHLKADVVISAFGSVLSDPKVKEA
                     LSPIKFNRWGLPEVDPETMQTSEAWVFAGGDVVGLANTTVESVNDGKQASWYIHKYVQ
                     SQYGASVSAKPELPLFYTPIDLVDISVEMAGLKFINPFGLASATPATSTSMIRRAFEA
                     GWGFALTKTFSLDKDIVTNVSPRIIRGTTSGPMYGPGQSSFLNIELISEKTAAYWCQS
                     VTELKADFPDNIVIASIMCSYNKNDWTELAKKSEDSGADALELNLSCPHGMGERGMGL
                     ACGQDPELVRNICRWVRQAVQIPFFAKLTPNVTDIVSIARAAKEGGANGVTATNTVSG
                     LMGLKSDGTPWPAVGIAKRTTYGGVSGTAIRPIALRAVTSIARALPGFPILATGGIDS
                     AESGLQFLHSGASVLQVCSAIQNQDFTVIEDYCTGLKALLYLKSIEELQDWDGQSPAT
                     VSHQKGKPVPRIAELMDKKLPSFGPYLEQRKKIIAENKIRLKEQNVAFSPLKRNCFIP
                     KRPIPTIKDVIGKALQYLGTFGELSNVEQVVAMIDEEMCINCGKCYMTCNDSGYQAIQ
                     FDPETHLPTITDTCTGCTLCLSVCPIVDCIKMVSRTTPYEPKRGVPLSVNPVC"
BASE COUNT     1317 a    845 c    949 g   1298 t
ORIGIN
```

```
  Query Match           19.7%;  Score 169.6;  DB 97;  Length 4409;
  Best Local Similarity  92.7%;  Pred. No. 8.5e-24;
  Matches  178;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy     244 CTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCG 303
           | | | | ||||     || || ||||||||||||||||||||||||||| |||||||||
Db    1817 CACCAAAACTTTCTCTCTTGATAAGGACATTGTGACAAATGTTTCCCCCAGAATCATCCG 1876

Qy     304 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1877 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 1936

Qy     364 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 423
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1937 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 1996

Qy     424 CCCAGACAACGT 435
           |||||||||  |
Db    1997 TCCAGACAACAT 2008
```

```
RESULT   8
BTU20981
LOCUS       BTU20981     4421 bp    mRNA            MAM       17-JAN-1997
DEFINITION  Bos taurus liver dihydropyrimidine dehydrogenase mRNA, complete
            cds.
ACCESSION   U20981
VERSION     U20981.1  GI:677950
KEYWORDS    .
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 4421)
  AUTHORS   Lu,Z., Zhang,R. and Diasio,R.B.
  TITLE     Dihydropyrimidine dehydrogenase activity in human peripheral blood
            mononuclear cells and liver: population characteristics, newly
            identified deficient patients, and clinical implication in
            5-fluorouracil chemotherapy
  JOURNAL   Cancer Res. 53 (22), 5433-5438 (1993)
  MEDLINE   94036837
REFERENCE   2  (bases 1 to 4421)
  AUTHORS   Albin,N., Johnson,M.R. and Diasio,R.B.
  TITLE     cDNA cloning of bovine liver dihydropyrimidine dehydrogenase
  JOURNAL   DNA Seq. 6 (4), 243-250 (1996)
  MEDLINE   97069948
REFERENCE   3  (bases 1 to 4421)
  AUTHORS   Johnson,M.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-FEB-1995) Pharmacology/Toxicology, University of
            Alabama at Birmingham, 1670 University Blvd., Birmingham, AL
            35294-0019, USA
FEATURES             Location/Qualifiers
     source          1. .4421
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /tissue_type="liver"
     CDS             75. .3152
                     /EC_number="1.3.1.2"
                     /note="dihydrothymine dehydrogenase; DPD"
                     /codon_start=1
                     /product="dihydropyrimidine dehydrogenase"
                     /protein_id="AAB40985.1"
                     /db_xref="GI:677951"
                     /translation="MAPVLSKDVADIESILALNPRTQSRATLRSTLAKKLDKKHWKRN
                     PDKNCFNCEKLENNFDDIKHTTLGERGALREAMRCLKCADAPCQKSCPTNLDIKSFIT
                     SISNKNYYGAAKMIFSDNPLGLTCGMVCPTSDLCVGGCNLYATEEGPINIGGLQQYAT
                     EVFKAMNIPQIRNPSLPPPEKMPEAYSAKIALLGAGPASISCASFLARLGYNDITIFE
                     KQEYVGGISTSEIPQFRLPYDVVNFEIELMKDLGVKIICGKSLSVNDITLSTLKEEGY
                     KAAFIGIGLPEPKKDHIFQGLTQDQGFYTSKDFLPLVAKSSKAGMCACHSPLLSIRGT
                     VIVLGAGDTAFDCATSALRCGARRVFIVFRKGFVNIRAVPEEVELAREEKCEFLPFLS
                     PRKVIVKGGRIVAMQFVRTEQDETGKWNEDGDQIACLKADVVISAFGSVLSDPKVKEA
                     LSPIKFNRWDLPEVDPETMQTSEPWVFAGGDVVGIANTTVEAVNDGKQASWYIHRYIQ
                     SQYGASVSAKPELPLFYTPIDLVDISVEMAALKFTNPFGLASATPTTSSSMIRRAFEA
                     GWAFALTKTFSLDKDIVTNVSPRIIRGTTSGPMYGPGQSSFLNIELISEKTAAYWCQS
                     VTELKADFPDNIVIASIMCSYNRNDWMELSRKAEASGADALELNLSCPHGMGERGMGL
                     ACGQDPELVRNICRWVRQAVRIPFFAKLTPNVTDIVSIARAAKEGGANGVTATNTVSG
                     LMGLKADGTPWPAVGREKRTTYGGVSGTAIRPIALRAVTTIARALPEFPILATGGIDS
                     AESGLQFLHGGASVLQVCSAIQNQDFTIIQDYCTGLKALLYLKSIEELQDWDGQSPAT
                     KSHQKGKPVPCIAELVGKKLPSFGPYLEKCKKIIAEEKLRLKKENVTVLPLERNHFIP
```

KKPIPSVKDVIGKALQYLGTYGELNNTEQVVAVIDEEMCINCGKCYMTCNDSGYQAIQ
FDPETHLPTVTDTCTGCTLCLSVCPIIDCIKMVSRTTPYEPKRGLPLAVNPVS"
BASE COUNT 1315 a 874 c 958 g 1274 t
ORIGIN

Query Match 17.8%; Score 153.4; DB 7; Length 4421;
Best Local Similarity 88.8%; Pred. No. 1.3e-20;
Matches 166; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```
Qy    246 CAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGG 305
          | | | ||||     || || |||| || ||||||||||||||| |||| ||||||||||
Db   1792 CCAAAACTTTCTCTCTTGATAAGGATATAGTGACAAATGTTTCACCCAGAATCATCCGGG 1851

Qy    306 GAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCA 365
          | |||||||||||||||||||||||||||||||||||||||| || ||||||||||||||||
Db   1852 GGACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCTTTCCTGAATATTGAGCTCA 1911

Qy    366 TCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCC 425
          ||||||| |||||||||||||||||||||||||||||||||||||||| || ||||| |
Db   1912 TCAGTGAAAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAAGCCGACTTTC 1971

Qy    426 CAGACAA 432
          |||||||
Db   1972 CAGACAA 1978
```

RESULT 9
SSU09179
LOCUS SSU09179 4441 bp mRNA MAM 28-DEC-1994
DEFINITION Sus scrofa dihydropyrimidine dehydrogenase mRNA, complete cds.
ACCESSION U09179
VERSION U09179.1 GI:558306
KEYWORDS .
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 4441)
AUTHORS Yokota,H., Fernandez-Salguero,P., Furuya,H., Lin,K., McBride,O.W.,
Podschun,B., Schnackerz,K.D. and Gonzalez,F.J.
TITLE cDNA cloning and chromosome mapping of human dihydropyrimidine
dehydrogenase, an enzyme associated with 5-fluorouracil toxicity
and congenital thymine uraciluria
JOURNAL J. Biol. Chem. 269 (37), 23192-23196 (1994)
MEDLINE 94365020
REFERENCE 2 (sites)
AUTHORS Eggink,G., Engel,H., Vriend,G., Terpstra,P. and Witholt,B.
TITLE Rubredoxin reductase of Pseudomonas oleovorans. Structural
relationship to other flavoprotein oxidoreductases based on one NAD
and two FAD fingerprints
JOURNAL J. Mol. Biol. 212 (1), 135-142 (1990)
MEDLINE 90204534
REFERENCE 3 (sites)
AUTHORS Porter,D.J., Chestnut,W.G., Merrill,B.M. and Spector,T.
TITLE Mechanism-based inactivation of dihydropyrimidine dehydrogenase by
5-ethynyluracil
JOURNAL J. Biol. Chem. 267 (8), 5236-5242 (1992)
MEDLINE 92184771
REFERENCE 4 (sites)
AUTHORS Dupuis,A., Skehel,J.M. and Walker,J.E.
TITLE A homologue of a nuclear-coded iron-sulfur protein subunit of
bovine mitochondrial complex I is encoded in chloroplast genomes
JOURNAL Biochemistry 30 (11), 2954-2960 (1991)
MEDLINE 91175743
REFERENCE 5 (sites)
AUTHORS Wierenga,R.K., De Maeyer,M.C.H. and Hol,W.G.J.
TITLE Interaction of pyrophosphatase moieties with alfa-helixes in
dinucleotide binding proteins
JOURNAL Biochemistry 24, 1346-1357 (1985)
REFERENCE 6 (bases 1 to 4441)
AUTHORS Gonzalez,F.
TITLE Direct Submission
JOURNAL Submitted (22-APR-1994) Frank J. Gonzalez, National Institutes of
Health, National Cancer Institute, 9000 Rockville Pike, Bethesda,
MD 20892, USA
FEATURES Location/Qualifiers
source 1. .4441
/organism="Sus scrofa"
/db_xref="taxon:9823"
/tissue_type="liver"
CDS 82. .3159
/codon_start=1
/product="dihydropyrimidine dehydrogenase"
/protein_id="AAA57475.1"
/db_xref="GI:558307"
/translation="MAPVLSKDVADIESILALNPRTQSHAALHSTLAKKLDKKHWKRN
PDKNCFHCEKLENNFGDIKHTTLGERGALREAMRCLKCADAPCQKSCPTHLDIKSFIT
SISNKNYYGAAKMIFSDNPLGLTCGMVCPTSDLCVGGCNLYATEEGSINIGGLQQFAS
EVFKAMNIPQIRNPCLPSQEKMPEAYSAKIALLGAGPASISCASFLARLGYSDITIFE
KQEYVGGLSTSEIPQFRLPYDVVNFEIELMKDLGVKIICGKSLSENEITLNTLKEEGY
KAAFIGIGLPEPKTDDIFQGLTQDQGFYTSKDFLPLVAKSSKAGMCACHSPLPSIRGA
VIVLGAGDTAFDCATSALRCGARRVFLVFRKGFVNIRAVPEEVELAKEEKCEFLPFLS
PRKVIVKGGRIVAVQFVRTEQDETGKWNEDEDQIVHLKADVVISAFGSVLRDPKVKEA
LSPIKFNRWDLPEVDPETMQTSEPWVFAGGDIVGMANTTVESVNDGKQASWYIHKYIQ
AQYGASVSAKPELPLFYTPVDLVDISVEMAGLKFINPFGLASAAPTTSSSMIRRAFEA
GWGFALTKTFSLDKDIVTNVSPRIVRGTTSGPMYGPGQSSFLNIELISEKTAAYWCQS
VTELKADFPDNIVIASIMCSYNKNDWMELSRKAEASGADALELNLSCPHGMGERGMGL
ACGQDPELVRNICRWVRQAVQIPFFAKLTPNVTDIVSIARAAKEGGADGVTATNTVSG
LMGLKADGTPWPAVGAGKRTTYGGVSGTAIRPIALRAVTTIARALPGFPILATGGIDS
AESGLQFLHSGASVLQVCSAVQNQDFTVIQDYCTGLKALLYLKSIEELQGWDGQSPGT
ESHQKGKPVPRIAELMGKKLPNFGPYLEQRKKIIAEEKMRLKEQNAAFPPLERKPFIP
KKPIPAIKDVIGKALQYLGTFGELSNIEQVVAVIDEEMCINCGKCYMTCNDSGYQAIQ
FDPETHLPTVTDTCTGCTLCLSVCPIIDCIRMVSRTTPYEPKRGLPLAVNPVC"
misc_feature 1084. .1134
/citation=[5]
/function="catalytic cofactor NADPH/NADP binding site"
misc_feature 1492. .1524
/citation=[2]
/function="electron transfer center, FAD binding site"
misc_feature 2062. .2175
/citation=[3]
/function="uracil (substrate) binding site"
misc_feature 2938. .2973
/standard_name="iron-sulfur center"
/citation=[4]
/function="catalytic cofactor [4Fe-4S] binding site"
BASE COUNT 1281 a 928 c 998 g 1234 t
ORIGIN

Query Match 17.4%; Score 150.2; DB 7; Length 4441;
Best Local Similarity 87.7%; Pred. No. 5.4e-20;
Matches 164; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

```
Qy    246 CAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGG 305
          | | | ||||     || || ||||||| |||||||||| || |||| |||| ||||||
Db   1799 CCAAAACTTTCTCTCTTGATAAGGACATAGTGACAAATGTCTCACCCAGAATCGTCCGGG 1858

Qy    306 GAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCA 365
          | || ||||||||||||||||| |||||||||||||||||||| ||||||||||||||||
Db   1859 GGACTACCTCTGGCCCCATGTACGGCCCTGGACAAAGCTCCTTCCTGAATATTGAGCTCA 1918

Qy    366 TCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCC 425
          ||||||| ||||| |||||||||||||||||||||||||||||||||| |||||||| |
Db   1919 TCAGTGAAAAAACAGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAAGCTGACTTTC 1978

Qy    426 CAGACAA 432
          |||||||
Db   1979 CAGACAA 1985
```

RESULT 10
AR027741
LOCUS AR027741 4447 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 3 from patent US 5856454.

```
ACCESSION   AR027741
VERSION     AR027741.1  GI:5938561
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4447)
  AUTHORS   Gonzalez,F.J. and Fernandez-Salguero,P.
  TITLE     CDNA for human and pig dihydropyrimidine dehydrogenase
  JOURNAL   Patent: US 5856454-A 3 05-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .4447
                     /organism="unknown"
BASE COUNT     1283 a    930 c   1000 g   1234 t
ORIGIN
```

Query Match 17.4%; Score 150.2; DB 9; Length 4447;
Best Local Similarity 87.7%; Pred. No. 5.4e-20;
Matches 164; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

```
Qy     246 CAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGG 305
           | | | ||||     || || ||||||| |||||||||| || |||| |||| ||||||
Db    1805 CCAAAACTTTCTCTCTTGATAAGGACATAGTGACAAATGTCTCACCCAGAATCGTCCGGG 1864

Qy     306 GAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCA 365
           | || ||||||||||||||||| |||||||||||||||||||| ||||||||||||||||
Db    1865 GGACTACCTCTGGCCCCATGTACGGCCCTGGACAAAGCTCCTTCCTGAATATTGAGCTCA 1924

Qy     366 TCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCC 425
           ||||||| ||||| |||||||||||||||||||||||||||||||||||||| |||||||| |
Db    1925 TCAGTGAAAAAACAGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAAGCTGACTTTC 1984

Qy     426 CAGACAA 432
           |||||||
Db    1985 CAGACAA 1991
```

```
RESULT  11
D85035
LOCUS       D85035       4358 bp    mRNA            ROD       19-SEP-1998
DEFINITION  Rattus norvegicus mRNA for dihydropyrimidine dehydrogenase,
            complete cds.
ACCESSION   D85035
VERSION     D85035.1  GI:3628592
KEYWORDS    DPD; dihydropyrimidine dehydrogenase.
SOURCE      Rattus norvegicus (strain:Wistar) 7-d-old male liver cDNA to mRNA,
            clone_lib:Lambda ZAPII cDNA.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 4358)
  AUTHORS   Kimura,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-1996) to the DDBJ/EMBL/GenBank databases.
            Masahiro Kimura, Kobe Gakuin University, Faculty of Nutrition; 518
            Arise, Ikawadani-cho, Nishi-ku, Kobe, Hyogo 651-21, Japan
            (Tel:078-974-1551(ex.3242), Fax:078-974-5689)
REFERENCE   2  (bases 1 to 4358)
  AUTHORS   Kimura,M., Sakata,S.F., Matoba,Y., Matsuda,K., Kontani,Y.,
            Kaneko,M. and Tamaki,N.
  TITLE     Cloning of rat dihydropyrimidine dehydrogenase and correlation of
            its mRNA increase in the rat liver with age
  JOURNAL   J. Nutr. Sci. Vitaminol. 44 (4), 537-546 (1998)
  MEDLINE   99037149
FEATURES             Location/Qualifiers
     source          1. .4358
                     /organism="Rattus norvegicus"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /clone_lib="Lambda ZAPII cDNA"
                     /dev_stage="7-d-old"
                     /sex="male"
                     /tissue_type="liver"
     gene            59. .3136
                     /gene="DPD"
     CDS             59. .3136
                     /gene="DPD"
                     /EC_number="1.3.1.2"
                     /codon_start=1
                     /product="dihydropyrimidine dehydrogenase"
                     /protein_id="BAA33218.1"
                     /db_xref="GI:3628593"
                     /translation="MAGVLSRDAPDIESILALNPRIQAHATLRSTMAKKLDKKHWKRN
                     TDKNCFICEKLENNFDDIKHTTLGERGALREAVRCLKCADAPCQKSCPTSLDIKSFIT
                     SIANKNYYGAAKLIFSDNPLGLTCGMVCPTSDLCVGGCNLHATEEGPINIGGLQQFAT
                     EVFKAMNIPQIRSPLLPPPEHMPEAYSAKIALFGAGPASISCASFLARLGYSDITIFE
                     KQEYVGGLSTSEIPQFRLPYDVVNFEIELMKDLGVKIICGKSISTDEMTLSTLKENGY
                     KAAFIGIGLPEPKKDHIFQGLTQVQGFYTSKDFLPLVAKGSKPGMCACHSPLPSVRGA
                     VIVLGAGDTAFDCATSALRCGARRVFIVFRKGFANIRAVPEEMELAKEEKCEFLPFLS
                     PRKVIVKDGKIVGMQFVRTEQDETGNWVEDEEQIVRLKADVVISPFGSVLDDPKVIEA
                     LSPIKFNRWGLPEVNPETMQTSEPWVFAGGDVVGMANTTVESVNDGKQASWYIHEYIQ
                     AQYGALVPSQPTLPLFYTPVDLVDISVEMAGLRFPNPFGLASATPATSTPMIRRAFEA
                     GWGFALTKTFSLDKDIVTNVSPRIIRGTTSGPLYGPGQSSFLNIELISEKTAAYWCHS
                     VTELKADFPDNILIASIMCSYNKNDWMELSKMAEASGADALELNLSCPHGMGERGMGL
                     ACGQDPELVRNICRWVRQSVRVPFFAKLTPNVTDIVSIARAAKEGGADGVTATNTVSG
                     LMGLKADGSPWPSVGSGKRTTYGGVSGTTIRPIALRAVTAIARALPGFPILATGGIDS
                     AESGLQFLHSGASVLQVCSAIQNQDFTVIEDYCTGLKALLYLKSIEELSDWDGQSPPT
                     MSHQKGKPVPHIAELMGQKLPSFGPYLERRKKILAASKIRENDQNRACSPLQRKHFNS
                     QKPIPAIKDVIGKSLQYLGTFGELNIMEQVVALIDEEMCINCGKCYMTCNDSGYQAIQ
                     FDPETHLPTVSDTCTGCTLCLSVCPIMDCIRMVSRATPYEPKRGLPLAVKPVC"
BASE COUNT     1234 a    930 c   1022 g   1172 t
ORIGIN
```

Query Match 16.9%; Score 145.2; DB 94; Length 4358;
Best Local Similarity 85.3%; Pred. No. 5.1e-19;
Matches 162; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

```
Qy     246 CAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGG 305
           | | | ||||     || || ||||||| |||||||| || || |||| ||||||||| |
Db    1776 CCAAAACTTTCTCTCTTGATAAGGACATCGTGACAAACGTCTCACCCAGAATCATCCGAG 1835

Qy     306 GAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCA 365
           | ||||| |||||||||| ||||||||||||||||||||||| || || || ||||||||||
Db    1836 GGACCACTTCTGGCCCCTTGTATGGCCCTGGACAAAGCTCTTTCCTCAACATTGAGCTCA 1895

Qy     366 TCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCC 425
           |||||||||||| |||||||||||||||| |||||||| |||||||||||||||||||||
Db    1896 TCAGTGAGAAAACAGCTGCATATTGGTGTCACAGTGTCACCGAACTAAAGGCTGACTTCC 1955

Qy     426 CAGACAACGT 435
           | |||||| |
Db    1956 CGGACAACAT 1965
```

```
RESULT  12
CELC25F6
LOCUS       CELC25F6    43803 bp    DNA             INV       02-NOV-1995
DEFINITION  Caenorhabditis elegans cosmid C25F6.
ACCESSION   U39742
VERSION     U39742.1  GI:1049455
KEYWORDS    .
SOURCE      Caenorhabditis elegans strain=Bristol N2.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 43803)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
```

```
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 43803)
  AUTHORS   Bentley,D.
  TITLE     The sequence of C. elegans cosmid C25F6
  JOURNAL   Unpublished (1995)
REFERENCE   3  (bases 1 to 43803)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-OCT-1995) Robert Waterston
COMMENT               Submitted by:
                      Genome Sequencing Center
                      Department of Genetics, Washington University,
                      St. Louis, MO  63110, USA, and
                      Sanger Centre, Hinxton Hall
                      Cambridge CB10 IRQ, England
                      e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk
                      NEIGHBORING COSMID INFORMATION:
            The 5' cosmid is C41A3, 1000 bp overlap; 3' end lies in a gap
            followed by the cosmid T23F2. Actual start of this cosmid is at
            base position 1 of CELC25F6; actual end is at base position 43803
            of CELC25F6.

            NOTES:
            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES              Location/Qualifiers
     source           1. .43803
                      /organism="Caenorhabditis elegans"
                      /db_xref="taxon:6239"
                      /map="X"
     tRNA             complement(1397. .1495)
                      /note="Leu;  codon recognized: CTA"
     gene             complement(5095. .5656)
                      /gene="C25F6.6"
     CDS              complement(join(5095. .5215,5259. .5385,5443. .5535,
                      5584. .5656))
                      /gene="C25F6.6"
                      /codon_start=1
                      /evidence=not_experimental
                      /protein_id="AAA80431.1"
                      /db_xref="GI:1049456"
                      /translation="MMISRFLSKILIVSCLIGVLGVADASSSMYARRRTASERLMSII
                      RPRMFQNPEFRSSQKIQKSIQLNPNRALPCCKDESGGSICKNLRRTDIKLFTQKCQTE
                      PDFSLVVCCSSCSDAGISYRKRAQKQVFIFTILVI"
     tRNA             complement(6056. .6128)
                      /note="Lys;  codon recognized: AAA"
     gene             16243. .20895
                      /gene="C25F6.4"
     CDS              join(16243. .16401,16449. .16601,17110. .17257,17309. .17421,
                      17576. .17675,17720. .17825,18125. .18202,18252. .18358,
                      18406. .18572,18930. .19243,19293. .19592,20318. .20513,
                      20569. .20718,20773. .20895)
                      /gene="C25F6.4"
                      /note="similar to protein tyrosine kinase"
                      /codon_start=1
                      /evidence=not_experimental
                      /protein_id="AAA80432.1"
                      /db_xref="GI:1049457"
                      /translation="MQNGDIADSQITASSSFDKQSVGPQNARLHSELASGAWCPKPQI
                      NSKSYEFLQVTLNDTFLITSVETQGRYGNGTGREFASHYMIDYLRPGSQWIRYKNRTG
                      HVYMDGNFDTTTPVIRVLDPPIVASRIRFVPSSKNTRTVCMRAEIHGCKHEGVTYYST
                      VPDGSRLDTLDFKDSMFEDSQIYTESGIKRYGSWPANRRFRSAASPFEKNQMNNSWIG
                      WNRDTTDGRITILFEFEEVHNFTDVVLATFGNRIDGIDVIFSQDGKTFPLFSQISSSE
                      RQSLNNTSRRYDFRVPLHNRAGRKVRISIKFSSDWMFLTEVHFTSAANLTLLSEKIPP
                      PSSAATQQLLVVCGIIFLTIFACVAYCVSVCLKRRQKNKSVDSNVKKDLIITHMGNKP
                      TCHVFPSNGKLSNGHYEVANDILYARSQKSTLLSVSSKSTFSCRAIPPTWTDFNFPPP
                      PEGREEHTYSQPVSPENSSNGSYKSVRKIQALKKYPSSALLIGKAIGEGKFTMIKECI
                      IFGGLKCAHKSTKEEDCVHGTRALGDEIACLLQCGRHPRIVELFGVDESYNLLLEHVE
                      YGCIRNFWMASEAPLDTEFLVRICKDIYSAMAYLESIRIVHGHFTPNNILMDGEFHAK
                      VCSPRGPSHHAQLRYSAPESIVNNEFTHKSDAWAVATTVYEMAYQCRQRPYEELTNEQ
                      IVDNACALLDHQPNAVVPLMPTVFNYEILQLLTRCFRVDQLERPTFERLVKPFQD"
     gene             22629. .27352
                      /gene="C25F6.3"
     CDS              join(22629. .22778,22996. .23109,23156. .23244,23886. .24135,
                      24185. .24653,24703. .24875,24923. .25050,25114. .25196,
                      25344. .25495,25544. .26161,26225. .26956,27003. .27136,
                      27190. .27352)
                      /gene="C25F6.3"
                      /note="coded for by C. elegans cDNA yk28e9.3; coded for by
                      C. elegans cDNA yk28e9.5; coded for by C. elegans cDNA
                      yk28h9.3; coded for by C. elegans cDNA yk28h9.5; coded for
                      by C. elegans cDNA yk40e4.5; coded for by C. elegans cDNA
                      yk5h3.5;  weakly similar to dihydroorotate dehydrogenase
                      and glutamate synthase"
                      /codon_start=1
                      /protein_id="AAA80433.1"
                      /db_xref="GI:1049458"
                      /translation="MSYDFYAFCLSIFIFCKFFNIIFTAMTPKPNTTSPTNNLPLLSK
                      DSPDIESLLILNPKVQDKANAVPSAVTKKNKHNWKRNEEKGCGSTCGESKLKNDFRDI
                      KHTTLSERGALKEAMRCLKCADAPCQKSCPTQLDVKSFITSISNKNYYGAARQILSDN
                      PLGLTCGMICPTSDLCVGSCNLQASEEGAINIGGLQQYACDVFKQMNVRQIVSKEVRE
                      NRNASHKEQVALIGCGPASISCASFLARLGYTDITIYEKRAYIGGLSSAEIPQFRLPY
                      DVVDFEIQLARDIGVQIETNRPLGKDGLTLAKLKEQGAAAVFIGIGNPEPKIDPLFEG
                      LTIENGFYTSKNYLPAVAAASKPGMCGCKRTPLPTMRGRVVVLGAGDTAMDCATSALR
                      CGASRVTIAFRKGFTGIRAVPEEMEAAKEEKCEFLPFSAPRKINVKDGRIVSIEFNKT
                      EQDDNGKWYEDEEQIVILKCDYVISAFGSTLKEDAVLSALQPCQLNKWGGIEVDSTTQ
                      QTSEKWVFAGGDVAGVAETTVESVNDGKIAAWNMHRYIQSLHGNQVSETPELPQFFTP
                      IDEVDISVDMCGVKFENPFGLASAPPTTSGPMCRRAFEQGWGFILTKTYGLDKDLVTN
                      VSPRIVRGSTSGPLYGPNQGSFMNIELISEKSCEYWLQCIRELKRDHPTKIVIASIMC
                      VYNKADWIELATKSEEAGADILELNLSCPHGMGEKGMGLACGQSPEIVKEICRWVRAC
                      VKIPFFPKMTPNITDVREIARAARDGGASGVTATNTVSSLMHMKADGNAWPAIGSTKR
                      TTYGGMSGSAIRPIAMKAVSSIANELDGFPIMATGGIESAETGLGFLMAGASVLQVCS
                      AVQNQDFTVVDDYCTGLKALLYLSGAESLKNWDGQSPPIEKHQKGKPILLQGQKKMPF
                      FGKYRDEREKLEAIKLSESNLLDTENYHFASRPDTQVSRVPTVEDVIGKALPRIGPYV
                      TLDNQEQKVAIIDDDMCINCGKCYMTCNDSGYQAITFDPVTHQPHVTEDDCTGCTLCY
                      SVCPIPECIEMVPRTGPWKAPKRGVKPSVEPGTPKVVKVDQRGRVILDTTGGMQ"
     gene             34301. .39725
                      /gene="C25F6.2"
     CDS              join(34301. .34613,35496. .35812,35889. .36107,36155. .36251,
                      36299. .36389,36443. .36794,36851. .37157,37216. .37684,
                      37743. .38266,38350. .38632,38680. .38801,38857. .38870,
                      38931. .38949,39228. .39253,39483. .39725)
                      /gene="C25F6.2"
                      /note="coded for by C. elegans cDNA yk25e5.3; coded for by
                      C. elegans cDNA yk25e5.5;  similar to repeat guanylate
                      kinase domain of D. melanogaster lethal(1) discs large-1
                      tumor suppressor protein (SP:DLG1_DROME, P31007) and R.
                      norvegicus postsynaptic density protein 95 (PSD-95)
                      (SP:PSD9_RAT, P31016)"
                      /codon_start=1
                      /protein_id="AAA80434.1"
                      /db_xref="GI:1049459"
                      /translation="MAACLCVCVCVCACVCRKFLFLSDKLCLIAAQFFLFVFNASFLP
                      SSVFRLLEALKPVVYFYFQSQKLVDQFNFFNFQRPASKENRSEWEERLTEEGSMSHES
                      SEKAHKAIENVEDYCQTLTRHGNEELRTNLERVITTFKSNLMHSLLDIHDLYEQTLLS
                      ERKSDAEKNMEVRRVIERLEGGPHSYNSRPAATTSTSNYNLSSTTPLISDLRDRGGFS
                      YLNGGGLGNGLGNGLGNGLLSSPYNSSSTHYLHERQRQTSHDGTWRETTTRTVDTPSG
                      LERRVVEHTGVIDDHGRKWELENIVLEKGHTGLGFSITGGMDQPTEDGDTSIYVTNII
                      EGGAALADGRMRKNDIITAVNNTNCENVKHEVAVNALKSSGNVVSLSLKRRKDEAFLP
                      IGGNFGGSTSYLRSGVTPSVSAGNLQHAIHSPSAPIHPPPPPPVHHGSLSQLSVGQYR
                      STRPNTSVIDLVKGARGLGFSIAGGQGNEHVKGDTDIYVTKIIEEGAAELDGRLRVGD
                      KILEVDHHSLINTTHENAVNVLKNTGNRVRLLIQQGTGAIFNDSASQQFMPTTPILRP
                      SSVQDYNRSQMGSQSHLSYGGPLNTSYSSQAPIAIPLEPRPVQLVKGQNGLGFNIVGG
                      EDNEPIYISFVLPGGVADLSGNVKTGDVLLEVNGVVLRNATHKEAAEALRNAGNPVYL
                      TLQYRPQEYQIFESKIEKLRNDVIAQSRMGTLSRKSEYVRALFDYDPSRENSVAPHRS
                      MGFNYGDILHIINSSDDEWWTARKVHENGEETAEGVIPSKKRVEKRERLRRKQVNFNS
                      GSQSLGRNSSTTGLENRRGSRSQLSFSRKFPFVKSTDRLNDLNEESSNVAEEPVWSYQ
                      AVEQQAINYVRPVIILGALKDRINDELVNRDPSKFSSCVPHTSRPPREGEVNGRDYYF
                      VNKHNMEEDVKNNLFIEAGQFQNNLYGTSIQSVRDVANQGRHCILDVSGNAIRRLQSN
```

```
                     ANIQPISIFIKPSSAQQILELDSQLATNRQDDRAMSGEEAQAQYSRCHRIEQTFGDLF
                     TQEISNTPIWVPQTLLLELRNLSFLLPSPSSPNSIHFLLEIVTVCDILNASFVSTNRK
                     SLSYSRLSTDEKLRLISIFWDFYYSESILFSSIGSKLFFFNYF"
     tRNA            39962. .40034
                     /note="Met;  codon recognized: ATG"
     tRNA            40305. .40377
                     /note="Met;  codon recognized: ATG"
     tRNA            40596. .40668
                     /note="Met;  codon recognized: ATG"
     gene            41327. .42523
                     /gene="C25F6.1"
     CDS             join(41327. .41484,42079. .42208,42260. .42523)
                     /gene="C25F6.1"
                     /note="coded for by C. elegans cDNA CEESH77F"
                     /codon_start=1
                     /protein_id="AAA80435.1"
                     /db_xref="GI:1049460"
                     /translation="MAKMKMTQEELVQAFNQYMPKRPVYPPLVHDDDVLAHSATESEG
                     SDEEEDDLTKRLKVTLAFSTDMMSECRTAWEDSTCKPPIQYLEDNIQKDVEYILEKLV
                     NLKIKPNNPIEWLGYCLLNMNIDPRKSKDHKNKPSELLRDFKIEALVNMVWNMEEGG
                     SGSISTLHKQAPSPTAGKRRNPS"
BASE COUNT    13561 a   8218 c   7586 g  14438 t
ORIGIN


  Query Match             8.2%;  Score 70.6;  DB 6;  Length 43803;
  Best Local Similarity   63.0%;  Pred. No. 0.00021;
  Matches  109;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy     260 TTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAACCACCTCTGGC 319
           ||  |  |||||  | || || ||||| || ||    ||  |||| ||| | |||  |||
Db   25784 TTGGACAAGGACTTGGTCACCAATGTCTCTCCACGTATTGTCCGAGGATCTACCAGTGGA 25843

Qy     320 CCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCAGTGAGAAAACG 379
           ||  | ||||| ||    ||| | || ||  ||||||||||||| |||   ||||||
Db   25844 CCACTTTATGGACCAAACCAAGGATCATTCATGAATATTGAGCTTATCTCAGAGAAATCC 25903

Qy     380 GCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAGACAA 432
             || ||| |||  |||| || | |  ||  | ||   ||||   |||   ||
Db   25904 TGTGAATACTGGCTTCAATGTATTAGGGAGTTGAAACGTGACCATCCAACTAA 25956


RESULT  13
AC014157
LOCUS       AC014157    24429 bp    DNA             HTG       16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC014157
VERSION     AC014157.1  GI:6437178
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 24429)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10214237 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .24429
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     6835 a   5227 c   5359 g   7008 t
ORIGIN


  Query Match             8.1%;  Score 70;  DB 63;  Length 24429;
  Best Local Similarity   63.9%;  Pred. No. 0.00027;
  Matches  106;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy     267 AGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAACCACCTCTGGCCCCATGT 326
           ||||  | || || ||||| || ||    ||| || |||| ||||| || |||  || ||
Db   14044 AGGATCTGGTCACGAATGTCTCGCCGCGCATCGTCAGGGGCACCACGTCGGGCTACAAGT 14103

Qy     327 ATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCAGTGAGAAAACGGCTGCAT 386
           |||| ||    ||  |||  || ||||| || || || |||   |||||   ||| |  |
Db   14104 ATGGACCGCAGCAGGGCTGTTTCCTGAACATCGAACTAATCTCGGAGAAGCGGGCCGAGT 14163

Qy     387 ATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAGACAA 432
           | |||    ||     |    ||||| |||  ||||||||| || ||
Db   14164 ACTGGCTGAAATCGATTGGAGAACTGAAGCGTGACTTCCCCGAGAA 14209


RESULT  14
AE003446/c
LOCUS       AE003446   303092 bp    DNA             INV       04-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386054 section 30
            of 35, complete sequence.
ACCESSION   AE003446 AE002566
VERSION     AE003446.2  GI:10728536
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 303092)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
```

```
REFERENCE   2  (bases 1 to 303092)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7290992.
FEATURES             Location/Qualifiers
     source          1. .303092
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="X"
     mRNA            join(<6691. .6699,8631. .>8972)
                     /gene="CG15364"
                     /product="CT35383"
                     /db_xref="FLYBASE:FBan0015364"
                     /db_xref="FLYBASE:FBgn0030075"
                     /evidence=not_experimental
     gene            <6691. .>8972
                     /gene="CG15364"
                     /map="8C7-8C9"
                     /db_xref="FLYBASE:FBan0015364"
                     /db_xref="FLYBASE:FBgn0030075"
                     /evidence=not_experimental
     CDS             join(6691. .6699,8631. .8972)
                     /gene="CG15364"
                     /note="CG15364 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0015364"
                     /db_xref="FLYBASE:FBgn0030075"
                     /evidence=not_experimental
                     /protein_id="AAF46431.1"
                     /db_xref="GI:7290993"
                     /translation="MYMSLRSSARYAKAKSKSAINLNAHLKKQQKLEEGHPEKKRQDK
                     KILRQSVSEDQPIYANVDEVIPIHLEPETGTEVDALMGNAQVSGPKNPGTQLQTTSSI
                     CIRLHCIPSNKIPN"
     mRNA            join(<17819. .19495,19597. .19682,19745. .19970,20028. .20777,
                     20838. .20948,21010. .21879)
                     /gene="CG4040"
                     /product="CT13364"
                     /db_xref="FLYBASE:FBan0004040"
                     /db_xref="FLYBASE:FBgn0030076"
     gene            <17819. .>21879
                     /gene="CG4040"
                     /map="8C8-8C9"
                     /db_xref="FLYBASE:FBan0004040"
                     /db_xref="FLYBASE:FBgn0030076"
     CDS             join(17819. .19495,19597. .19682,19745. .19970,20028. .20777,
                     20838. .20948,21010. .21837)
                     /gene="CG4040"
                     /note="CG4040 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0004040"
                     /db_xref="FLYBASE:FBgn0030076"
                     /protein_id="AAF46432.1"
                     /db_xref="GI:7290994"
                     /translation="MNHLRVHKFKLGGKDNAAMADLPQSENTTNNNNNIGSNNNQKGG
                     AKGFLGRIKRYSVLGNKLGRRNLGSMNLNPPNNVVDNSVSPTKLHTGSYNMTGSHSDD
                     HRLEIGAPILISTTTLDTDRFDVTEARLKQIGGGIAQTSSIVRTLTPRSSDEEEFVDA
                     RSSPQQMERDEDLQIADDQVEKNDRDQYGHQFQLPKKQHREEEMGDLKYLDDLEEPEQ
                     FETPKHEVEEKLLTANQDVKRPITPITIAIDPPIRRPRIARQQRQSQSVQNLHRSELK
                     VYLHKSNSVALDMNVTAPAKLGLDLNLPELPELYGASELSLAASDNKENLPAVKPSAE
                     RSPIPNGGFMAQQPHFKSLDSFHLNTRSHSNQSSQQSSRQSLACSSSNGEYDFDLKSV
                     SYQSLNAQNLFVSIDELQELTRQINETEDFGKEIDLEYCAHRDQLRPSERRITLLKNK
                     NQTLINFNHNKEKLRKGWHGMKHWLGEEGSKIKEAVRQQTPLKRLAQSRSNLNQSTAD
                     ASRQSMSPERTQRDVTESCEDVTERTEMESSMSHRNSEEDLSPHAKRFKDEGQNGFEE
                     LRRYVKQGGDFSKELIFVLQERADSELIYSKSLSKLANKLNKAGREIPGSVADAWRGV
                     ATEMESRSDIHRQLAASLTDELVKPLKIVVEGHHKARKAVESNVDKAARVLGEWRASE
                     AKAKKASHTAARENEKLQDAMLDVRIQKSPSIALLHQGPNKQAAEKELKSAEKDCVKL
                     DNKRKKAEEAVKRADVEYYTLCVRSERARVDWEMAVLRGSAQLQSSEQQRLGNMHNFA
                     QQYARLISDMNPILGGLSTRLQPQLDACNVAKDMQVVRHIRRNSEGPSEQLLPDFYCE
                     HTTLAMNRERRKHALIKLLQLVKTDLERERRSRDGLRGLSQSLNHQEHQNITDKLYHI
                     RSMLTYLEGARLKLHSALLELDHKPRATHPLAQHIHITRDRTGLQQSILKVPNWLKNN
                     DKSQTQQSTSMLAHDITDITTNHEDELPDENCSHLHSSSHSNNNSNGSCTDVSSVVKP
                     FNRSKSNIETNFTSQPKIISTTNAAIAVSVKSKTLANLQDGHHNNQQHHHHHHNHHSD
                     RGQADGGSNQQDSDFDEFSSQDEDDEPQQPQQQPQPLQSQQQQQQPLLQNPTINGQVQ
                     TQHFYQNAQELQKGQKNGSVPILGRCKALYSYTPKLYDELELSPGDIIEVHAKQDDGW
                     WLGALRNHIGIFPATYVEECA"
     mRNA            complement(join(<23607. .24527,24584. .24901,25007. .25148,
                     25214. .25346,25568. .25910,26018. .>26626))
                     /gene="CG15365"
                     /product="CT35384"
                     /db_xref="FLYBASE:FBan0015365"
                     /db_xref="FLYBASE:FBgn0030077"
                     /evidence=not_experimental
     gene            complement(<23607. .>26626)
                     /gene="CG15365"
                     /map="8C9-8C10"
                     /db_xref="FLYBASE:FBan0015365"
                     /db_xref="FLYBASE:FBgn0030077"
                     /evidence=not_experimental
     CDS             complement(join(23607. .24527,24584. .24901,25007. .25148,
                     25214. .25346,25568. .25910,26018. .26626))
                     /gene="CG15365"
                     /note="CG15365 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0015365"
                     /db_xref="FLYBASE:FBgn0030077"
                     /evidence=not_experimental
                     /protein_id="AAF46433.1"
                     /db_xref="GI:7290995"
                     /translation="MSAATHMAMPAHAVVPPPGTPPVLGGSGCSNLSLPSGKRGHKRS
                     VSLENNAPLALRSTPSMGQSSLGSPLLQFGQAAPGQGHSQGFHTGTLKSRQEQRRLSQ
                     KFGSHSNLDDEGVGIGVGMGMGMGLQMRSKFQNIRQMFELSRSCVGGGGGDGRGGAAE
                     EEAMQSLPATLPQQHQQQQQRRTTPIAGGSRHQQQQQMGEAEQTSGNFLRPIAFKPIP
                     FEPDYRIACQQQQQHQQQLQLQQQQQQQQQQQQQQQQQQQLQLQHQQQQQHQQQHQHQHQQQ
                     HQSLQLPTEGGQAGGVAAERYGYGSTPSLAPLPTASAQKFGSTTDLRHLAARRRNHRL
                     LQRCSNEDALTLDLLSTGSHDRPDGASSKSSGTVVGSSDLTPSPSDSGISELEAALKD
                     RDSELSYLRQAMEHNEKDKEVYWEHETQRLRLFYEGQQRECQLKLRKMEQLLGLQQFQ
                     LKQHKLRQSEQVNRLQQQLDLARDSSQGLQQQVDALRHQLEDSEWRVCERNGEIALLK
                     TQLKEAHLEINMKDHAIVSLKHSSNRSSSSSNSNSNKSCDTNTSQSQPKQQAKEEQQQ
                     QQHPELQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQLKQQHQPDQQQLQKI
                     ITLKDQVIGALTSELAKLRKELSDLAIAHEYGEEPCGRYTRLKQQLDNLNEICQKTRK
                     YTTTSASPTAVASPAAQVAPPAPSPAAHVATPSCAQELPKLDVLVQRLQLDQGQGTQQ
                     TLDTLIEESTTYAEDIAKLRQKLDDFRLNLELEKRQWSAEKDKVLGYQKQLQAHYIQM
                     YQKLRCLDSAGAAVGQVEATSGN"
     mRNA            complement(38162. .38842)
                     /gene="CG10970"
                     /product="CT30739"
                     /db_xref="FLYBASE:FBan0010970"
                     /db_xref="FLYBASE:FBgn0030078"
     gene            complement(<38162. .>38842)
                     /gene="CG10970"
                     /map="8C9-8C12"
                     /db_xref="FLYBASE:FBan0010970"
                     /db_xref="FLYBASE:FBgn0030078"
     CDS             complement(38314. .38625)
                     /gene="CG10970"
                     /note="CG10970 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0010970"
                     /db_xref="FLYBASE:FBgn0030078"
                     /protein_id="AAF46434.1"
                     /db_xref="GI:7290996"
                     /translation="MAGQKEEILCCDFEVRGRVPKDAFELFALVQANALGLRGFIAQV
                     SEEKFKGRLEGEGKVIDAFKKLILAAADYVQAIKEFIIKNLKVIQEYTYKTFEIKVNL
                     K"
     mRNA            complement(join(<40014. .40361,40432. .40549,41242. .41250))
                     /gene="CG7267"
                     /product="CT22403"
                     /db_xref="FLYBASE:FBan0007267"
                     /db_xref="FLYBASE:FBgn0030079"
     gene            complement(<40014. .>41250)
                     /gene="CG7267"
                     /map="8C9-8C12"
                     /db_xref="FLYBASE:FBan0007267"
                     /db_xref="FLYBASE:FBgn0030079"
```

```
     CDS             complement(join(40144. .40361,40432. .40549,41242. .41250))

  Query Match             8.1%;  Score 70;  DB 4;  Length 303092;
  Best Local Similarity  63.9%;  Pred. No. 0.00027;
  Matches  106;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy     267 AGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAACCACCTCTGGCCCCATGT 326
           ||||  | || || ||||| || ||    ||| || |||| ||||| || |||  || ||
Db  165897 AGGATCTGGTCACGAATGTCTCGCCGCGCATCGTCAGGGGCACCACGTCGGGCTACAAGT 165838

Qy     327 ATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCAGTGAGAAAACGGCTGCAT 386
           |||| ||    ||  |||  || ||||| || || || |||   |||||   ||| |  |
Db  165837 ATGGACCGCAGCAGGGCTGTTTCCTGAACATCGAACTAATCTCGGAGAAGCGGGCCGAGT 165778

Qy     387 ATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAGACAA 432
           | |||    ||    |    ||||| |||  ||||||||| || ||
Db  165777 ACTGGCTGAAATCGATTGGAGAACTGAAGCGTGACTTCCCCGAGAA 165732


RESULT  15
DMU65491
LOCUS       DMU65491     2348 bp    mRNA            INV       26-SEP-1996
DEFINITION  Drosophila melanogaster Dreg-3 protein mRNA, complete cds.
ACCESSION   U65491
VERSION     U65491.1  GI:1561729
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 2348)
  AUTHORS   Van Gelder,R.N., Bae,H., Palazzolo,M.J. and Krasnow,M.A.
  TITLE     Extent and character of circadian gene expression in Drosophila
            melanogaster: identification of twenty oscillating mRNAs in the fly
            head
  JOURNAL   Curr. Biol. 5 (12), 1424-1436 (1995)
  MEDLINE   96362140
REFERENCE   2  (bases 1 to 2348)
  AUTHORS   Van Gelder,R.N., Bae,H., Palazzolo,M.J. and Krasnow,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1996) Ophthalmology and Visual Sciences,
            Washington University School of Medicine, 660 S. Euclid Avenue, Box
            8096, St. Louis, MO 63110, USA
FEATURES             Location/Qualifiers
     source          1. .2348
                     /organism="Drosophila melanogaster"
                     /strain="Canton-S"
                     /db_xref="taxon:7227"
     CDS             358. .2238
                     /note="similar to dihydropyrimidine dehydrogenase;
                     rhythmically expressed gene 3"
                     /codon_start=1
                     /product="Dreg-3 protein"
                     /protein_id="AAC47288.1"
                     /db_xref="GI:1561730"
                     /translation="MEFCRTEQNENDEWVEDEEQTQRLKANFVISAFGSGLEDQDVKA
                     ALAPLQFRGELPVVDRVTMQSSVKQVFLGGDLAGVANTTVESVNDGKVAAWSIHCQLQ
                     GLPLDTPAALPLFYTDIDAVDISVEMCGIRFENPFGLASAPPTTSTAMIRRAFEQGWG
                     FVVTKTFGLDKDLVTNVSPRIVRGTTSGYKYGPQQGCFLNIELISEKRAEYWLKSIGE
                     LKRDFPEKIVIASIMCSFNEEDWTELAIKAEQSGADALELNLSCPHGMGERGMGLACG
                     QDPELVEQISRWVRKAVKLPFFIKLTPNITDIVSIAAAAKREEPMRSAINTVQGLMGL
                     KADSTAWPAIGKEQRTTYGGVSGNATRPMALKAISDIANRVPGFPILGIGGIDSGEVA
                     LQFIHAGATVLQICSSVQNQDFTVIEDYCTALKALLYLKRIRHQSMVPSGMASHHPRR
                     SIRRQARCPFDRRGKATLGFFGPYQRQRDIKMAELRSQKGALSWDAEQVKATPPASNG
                     APNPAPRIKDVIGAALDKIGSYNKLDNKQQKVALIDDDMCINCGKCYMTCADSGYQAI
                     EFDKDTHIPHVNDDCTGCTLCVSVCPIIDCITMVPKKIPHVIKRGVEEKIFYTHALSQ
                     CQ"
BASE COUNT      538 a    649 c    684 g    477 t
ORIGIN


  Query Match             7.9%;  Score 68.4;  DB 6;  Length 2348;
  Best Local Similarity  63.3%;  Pred. No. 0.00056;
  Matches  105;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy     267 AGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAACCACCTCTGGCCCCATGT 326
           ||||  | || || ||||| || ||    ||| || |||| ||||| || |||  || ||
Db     869 AGGATCTGGTCACGAATGTCTCGCCGCGCATCGTCAGGGGCACCACGTCGGGCTACAAGT 928

Qy     327 ATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCAGTGAGAAAACGGCTGCAT 386
           |||| ||    ||  | |  || ||||| || || || |||   |||||   ||| |  |
Db     929 ATGGACCGCAGCAGGGTTGTTTCCTGAACATCGAACTAATCTCGGAGAAGCGGGCCGAGT 988

Qy     387 ATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAGACAA 432
           | |||    ||    |    ||||| |||  ||||||||| || ||
Db     989 ACTGGCTGAAATCGATTGGAGAACTGAAGCGTGACTTCCCCGAGAA 1034


Search completed: November 20, 2001, 16:01:12
Job time: 5611 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2001, 14:29:36 ; Search time 123.96 Seconds
(without alignments)
4361.272 Million cell updates/sec

Title: US-09-308-080-1
Perfect score: 861
Sequence: 1 TGTTAATGAAGATAAATATT.....AGTGGGAAATAATTTATTAA 861

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601:*
1: /SIDS8/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	857.8	99.6	861	18	AAT91798	Dihydropyrimidine
2	171.2	19.9	3951	20	AAX22903	Human DPD DNA. Ho
3	169.6	19.7	3957	17	AAT14077	Human dihydropyrim
4	169.6	19.7	4368	16	AAT03133	Human dihydropyrim
5	169.6	19.7	4369	16	AAT03143	Human dihydropyrim
6	153.4	17.8	4414	16	AAT03132	Bovine dihydropyri
7	153.4	17.8	4430	16	AAT03142	Bovine dihydropyri
8	151.4	17.6	936	22	AAF58252	Oligonucleotide D1
9	151.4	17.6	936	22	AAF58254	Oligonucleotide D1
10	151.4	17.6	936	22	AAF58257	Oligonucleotide D1
11	151.4	17.6	936	22	AAF58259	Oligonucleotide D2
12	151.4	17.6	936	22	AAF58262	Oligonucleotide D2
13	151.4	17.6	938	22	AAF58255	Oligonucleotide D1
14	150.2	17.4	4447	17	AAT14078	Pig dihydropyrimid
c 15	147.4	17.1	936	22	AAF58252	Oligonucleotide D1
c 16	147.4	17.1	936	22	AAF58254	Oligonucleotide D1
c 17	147.4	17.1	936	22	AAF58257	Oligonucleotide D1
c 18	147.4	17.1	936	22	AAF58259	Oligonucleotide D2
c 19	147.4	17.1	936	22	AAF58262	Oligonucleotide D2
c 20	147.4	17.1	938	22	AAF58255	Oligonucleotide D1
21	61.2	7.1	244	22	AAF58238	Oligonucleotide D1
c 22	59.8	6.9	244	22	AAF58238	Oligonucleotide D1
c 23	45	5.2	5852	12	AAQ11710	Dictyostelium plas
24	42.4	4.9	3648	22	AAD02172	Plasmid pHiNk181 c
25	42.4	4.9	3648	22	AAD02173	Plasmid pHiNk184 c
26	42.4	4.9	11461	22	AAD02175	Plasmid pZU634 for
27	42.4	4.9	12766	22	AAD02174	Plasmid pZU623 for
28	42.4	4.9	13737	21	AAA54212	Transformation vec
c 29	42.2	4.9	2400	18	AAT49364	Nicotiana plumbagi
c 30	42.2	4.9	8318	20	AAX20264	Borrelia burgdorfe
c 31	41.4	4.8	745	20	AAX39751	Gastric cancer ass
c 32	41	4.8	3947	19	AAV69429	Ehrlichia sp. exte
c 33	41	4.8	3947	21	AAZ89994	Ehrlichia antigen
34	40.2	4.7	2313	21	AAC76449	Human ORFX ORF2004
35	40	4.6	37808	20	AAX02780	Vector pMVX-BG DNA
c 36	39.4	4.6	1717	20	AAX20311	Borrelia burgdorfe
c 37	39	4.5	14752	20	AAX20256	Borrelia burgdorfe
c 38	39	4.5	116277	20	AAX20249	Borrelia burgdorfe
39	39	4.5	910715	20	AAX20248	Borrelia burgdorfe
c 40	38.8	4.5	5810	18	AAT91323	Arabidopsis thalia
41	38.8	4.5	8911	20	AAX13588	Enterococcus faeca
42	38.6	4.5	1400	20	AAZ27735	Human DNA marker c
43	38.6	4.5	1591	21	AAZ95204	Human UGT2B7 exon
c 44	38.6	4.5	111309	20	AAX20250	Borrelia burgdorfe
45	38.4	4.5	1776	18	AAT91744	Tobacco calcium/ca

ALIGNMENTS

RESULT 1
AAT91798
ID AAT91798 standard; DNA; 861 BP.
XX
AC AAT91798;
XX
DT 14-APR-1998 (first entry)
XX
DE Dihydropyrimidine dehydrogenase genomic fragment.
XX
KW Human; dihydropyrimidine dehydrogenase; DPD; slicing defect;
KW detection; 5-fluorouracil; cancer; anticancer; uraciluria; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT exon 269..433
FT /*tag= a
FT /note= "Encodes amino acids 581-635 of the DPD protein"
XX
PN WO9735034-A1.
XX
PD 25-SEP-1997.
XX
PF 19-MAR-1997; 97WO-US04269.
XX
PR 20-MAR-1996; 96US-0013835.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Fernandez-Salguero P, Gonzalez FJ;
XX
DR WPI; 1997-480236/44.
DR P-PSDB; AAW30092.

```
XX
PT   Detecting a splicing defect in the di:hydro:pyrimidine dehydrogenase
PT   gene - used to identify subjects sensitive to 5-fluorouracil, toxic
PT   to individuals with DPD defects
XX
PS   Disclosure; Fig 1; 38pp; English.
XX
CC   A novel method has been developed for detecting the splicing defect in
CC   the dihydropyrimidine dehydrogenase (DPD) gene. The method comprises
CC   determining whether genomic DNA containing the DPD gene has a wild-type
CC   intron-exon boundary for an exon that encodes amino acids (aa) 581-635
CC   of the corresponding DPD protein. The present sequence represents a
CC   DPD genomic fragment which encodes the amino acids 581-635. The method
CC   is used specifically to determine sensitivity of subjects to the
CC   anticancer agent 5-fluorouracil, which is dangerously toxic to those
CC   with DPD defects. It can also be used to diagnose DPD-deficiency
CC   disorders such as uraciluria.
XX
SQ   Sequence 861 BP; 300 A; 140 C; 161 G; 260 T; 0 other;


  Query Match             99.6%;  Score 857.8;  DB 18;  Length 861;
  Best Local Similarity   99.8%;  Pred. No. 7.1e-183;
  Matches  859;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 TGTTAATGAAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 tgttaatgaagataaatatttttgttttttcgctgttctaaacctagggttacaagaagt 60

Qy      61 AATTTATCTGGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 aatttatctggagctaacaaatactttatttttacctttttatttgcaagtagtttatgtt 120

Qy     121 CAATTCTAATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 caattctaatttaatgtatattaaaaattcctctgcaaatatgtgaggagggacctcata 180

Qy     181 AAATATTGTCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 aaatattgtcatatggaaatgagcagataataaagattatagcttttctttgtcaaaagg 240

Qy     241 AGACTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCAT 300
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     241 agactcaatatctttactctttcatgaggacattgtgacaaatgtttcccccagaatcat 300

Qy     301 CCGGGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ccggggaaccacctctggccccatgtatggccctggacaaagctcctttctgaatattga 360

Qy     361 GCTCATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 gctcatcagtgagaaaacggctgcatattggtgtcaaagtgtcactgaactaaaggctga 420

Qy     421 CTTCCCAGACAACGTAAGTGTGATAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTG 480
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     421 cttcccagacaacggaagtgtgataaaaatctaaaacaagagaattggcataagttggtg 480

Qy     481 AATGTTTATTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCAA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 aatgtttatttaaacatccaattcataggcttataaatattaatgtgtatattttatcaa 540

Qy     541 CGAATCTGCCAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 cgaatctgccagttgctttgctgatgcatagaaagataaaaaagaaagaaaagctcaaga 600

Qy     601 ACTCATAAAAACCCACACAATGTGAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAA 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 actcataaaaacccacacaatgtgaagctctgttataaatgggtgccatgtaagatggaa 660

Qy     661 GAAGTATCTACATAAGCAGAAGGAAGAGAAATGAAATACTCATTTTATTGAGTTGGCCCC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 gaagtatctacataagcagaaggaagagaaatgaaatactcattttattgagttggcccc 720
```

```
Qy     721 CACTGTATGTGGCTGGTATTTATGAAGGTGATGACCCAGGAAGAAATTGTAAACTATAAA 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 cactgtatgtggctggtatttatgaaggtgatgacccaggaagaaattgtaaactataaa 780

Qy     781 CCACTCCAAATATAAACCCGAGGCAGAAGCAGCATATCTCCTATGAAGCCTGTATTTACT 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 ccactccaaatataaacccgaggcagaagcagcatatctcctatgaagcctgtatttact 840

Qy     841 CAGTGGGAAATAATTTATTAA 861
           |||||||||||||||||||||
Db     841 cagtgggaaataatttattaa 861


RESULT   2
AAX22903
ID   AAX22903 standard; DNA; 3951 BP.
XX
AC   AAX22903;
XX
DT   28-MAY-1999  (first entry)
XX
DE   Human DPD DNA.
XX
KW   DPD; dihydropyrimidine-dehydrogenase; monoclonal antibody; MAb;
KW   immunoassay reagent; cancer patient; treatment; antitumour agent;
KW   5-fluorouracil; affinity purification; toxicity; ss.
XX
OS   Homo sapiens.
XX
PN   DE19837391-A1.
XX
PD   25-FEB-1999.
XX
PF   18-AUG-1998;   98DE-1037391.
XX
PR   22-AUG-1997;   97EP-0114630.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI   Hasegawa M,  Yoshikubo T;
XX
DR   WPI; 1999-155202/14.
DR   P-PSDB; AAW93361.
XX
PT   Monoclonal antibody specific for dihydropyrimidine dehydrogenase -
PT   for assessing patient response to 5-fluorouracil antitumor agents
XX
PS   Disclosure; Page 15-17; 34pp; German.
XX
CC   This invention describes a monoclonal antibody (MAb) specific for
CC   dehydropyrimidine dehydrogenase (DPD). This MAb is used as immunoassay
CC   reagents to identify a lack of DPD in a patient and to assess the
CC   sensitivity of cancer patients to treatment with antitumor agents of the
CC   5-fluorouracil (5-FU) type. The MAb can also be used for affinity
CC   purification of DPD. DPD is involved in reduction of 5-FU (and related
CC   catabolites and derivatives) and lack of it is associated with increased
CC   toxicity of this type of antitumor agent. It has specific binding
CC   interaction. The MAb provide a sensitive and reliable test for DPD,
CC   which is simple, rapid and suitable for routine screening.
XX
SQ   Sequence 3951 BP; 1153 A; 785 C; 896 G; 1117 T; 0 other;


  Query Match             19.9%;  Score 171.2;  DB 20;  Length 3951;
  Best Local Similarity   93.2%;  Pred. No. 2.4e-29;
  Matches  179;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy     244 CTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCG 303
           | | | | ||||     || || |||||||||||||||||||||||||| |||||||||
Db    1797 caccaaaactttctctcttgataaggacattgtgacaaatgtttcccccagaatcatccg 1856

Qy     304 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1857 gggaaccacctctggccccatgtatggccctggacaaagctcctttctgaatattgagct 1916

Qy    364 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 423
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1917 catcagtgagaaaacggctgcatattggtgtcaaagtgtcactgaactaaaggctgactt 1976

Qy    424 CCCAGACAACGT 435
          |||||||||| |
Db   1977 cccagacaacat 1988


RESULT   3
AAT14077
ID   AAT14077 standard; cDNA; 3957 BP.
XX
AC   AAT14077;
XX
DT   11-JUN-1996  (first entry)
XX
DE   Human dihydropyrimidine dehydrogenase cDNA.
XX
KW   Dihydropyrimidine dehydrogenase; DPD; DPYD gene; probe; deficiency;
KW   5-fluorouracil; cytostatic; cancer; selectable marker; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             88..3165
FT                   /*tag= a
FT   primer_bind     complement (36..55)
FT                   /*tag= b
FT                   /note= "primer RTF1"
FT   primer_bind     1536..1558
FT                   /*tag= c
FT                   /note= "primer RTR1"
FT   primer_bind     complement (1539..1558)
FT                   /*tag= d
FT                   /note= "primer H13"
FT   primer_bind     2426..2445
FT                   /*tag= e
FT                   /note= "primer RTR4"
FT   primer_bind     complement (2424..2447)
FT                   /*tag= f
FT                   /note= "primer RTR5"
FT   primer_bind     3320..3343
FT                   /*tag= g
FT                   /note= "primer RTR5"
XX
PN   WO9608568-A2.
XX
PD   21-MAR-1996.
XX
PF   07-SEP-1995;   95WO-US12016.
XX
PR   12-SEP-1994;   94US-0304309.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Fernandez-Salguero P,  Gonzalez FJ;
XX
DR   WPI; 1996-179943/18.
DR   P-PSDB; AAR91420.
XX
PT   Pig and human di:hydro:pyrimidine dehydrogenase (DPD) genes and
PT   probes  - useful for detection of DPD deficiencies and
PT   identification of humans at risk of toxic reaction to
PT   5-fluoro:uracil anti-cancer treatment
XX
PS   Claim 1; Page 43-48; 78pp; English.
XX
CC   Human dihydropyrmidine dehydrogenase (DPD) cDNA (AAT14077) was
CC   isolated as 3 overlapping fragments obtd. from human liver
CC   cDNA libraries screened using pig DPD cDNA (see also AAT14077).
CC   The DPD gene, DPYD, is located at chromosome 1p22.  The cDNA
CC   can be inserted into a vector and used for prodn. of DPD (AAR91420)
CC   in transformed host cells, or used as a selectable marker that
CC   functions in prokaryotic and eukaryotic cells.  It can also be
CC   utilised as a probe to detect DPD deficiency in an individual,
CC   i.e. to identify risk of toxic reaction to 4-fluorouracil.
XX
SQ   Sequence 3957 BP; 1156 A; 787 C; 897 G; 1117 T; 0 other;


  Query Match             19.7%;  Score 169.6;  DB 17;  Length 3957;
  Best Local Similarity   92.7%;  Pred. No. 5.4e-29;
  Matches  178;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy    244 CTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCG 303
          | | | | ||||     || || ||||||||||||||||||||||||||| |||||||||
Db   1803 caccaaaactttctctcttgataaggacattgtgacaaatgtttcccccagaatcatccg 1862

Qy    304 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 363
          ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   1863 gggaaccacctctggccccatgtatgcccctggacaaagctcctttctgaatattgagct 1922

Qy    364 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 423
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1923 catcagtgagaaaacggctgcatattggtgtcaaagtgtcactgaactaaaggctgactt 1982

Qy    424 CCCAGACAACGT 435
          |||||||||| |
Db   1983 cccagacaacat 1994


RESULT   4
AAT03133
ID   AAT03133 standard; DNA; 4368 BP.
XX
AC   AAT03133;
XX
DT   04-JUN-1996  (first entry)
XX
DE   Human dihydropyrimidine dehydrogenase gene.
XX
KW   Bovine; liver; human; dihydropyrimidine dehydrogenase; DPD;
KW   pyrimidine catabolism; 5,6-dihydropyrimidine; pyrimidine analogue;
KW   fluoropyrimidine; anticancer drug; 5-fluorouracil; FUra; cancer;
KW   frameshift mutation; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             49..3126
FT                   /*tag= a
FT                   /product= DPD
XX
PN   WO9528489-A1.
XX
PD   26-OCT-1995.
XX
PF   13-APR-1995;   95WO-US04567.
XX
PR   13-APR-1994;   94US-0227357.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Cheng X,  Diasio RB,  Johnson M,  Lu Z,  Zhang R;
XX
DR   WPI; 1995-373803/48.
DR   P-PSDB; AAR86001.
XX
PT   Novel di:hydro:pyrimidine dehydrogenase gene - used to optimise
PT   5-fluoro:uracil doses given to cancer patients
XX
PS   Claim 10; Page 140-56; 207pp; English.
XX
```

```
CC   This sequence represents the DNA sequence encoding human
CC   dihydropyrimidine dehydrogenase (DPD).  DPD catalyses the initial
CC   and rate limiting step in pyrimidine catabolism, the reduction of
CC   pyrimidines to 5,6-dihydropyrimidines.  DPD is a complex enzyme
CC   consisting of two identical subunits, containing FMN, FAD and iron-
CC   sulphur centers, and utilising NADPH as a cofactor.  DPD has also been
CC   shown to catalyse the reduction of various pyrimidine analogues
CC   including the fluoropyrimidine anticancer drug 5-fluorouracil (FUra).
CC   Up to 85% of administered FUra may be catabolised by DPD, and it
CC   therefore governs the effectiveness of FUra as an anticancer drug.
CC   DPD genes or fragments of them may be used in the detection of DPD in a
CC   sample, esp. isolated from a cancer patient.  According to the amount
CC   of DPD detected, a therapeutically effective amount of FUra may be
CC   determined and administered.  DPD deficiency, leading to life-
CC   threatening toxicity on exposure to FUra, in a human caused by a
CC   frameshift mutation may be determined by means of a molecular biological
CC   assay to detect the deletion of an A residue at codon 318 within the
CC   DPD-coding region.
XX
SQ   Sequence 4368 BP; 1325 A; 829 C; 932 G; 1282 T; 0 other;


  Query Match             19.7%;  Score 169.6;  DB 16;  Length 4368;
  Best Local Similarity   92.7%;  Pred. No. 5.5e-29;
  Matches  178;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy     244 CTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCG 303
           | | | | ||||     || || ||||||||||||||||||||||||||| |||||||||
Db    1764 caccaaaactttctctcttgataaggacattgtgacaaatgtttcccccagaatcatccg 1823

Qy     304 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1824 gggaaccacctctggccccatgtatggccctggacaaagctcctttctgaatattgagct 1883

Qy     364 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 423
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1884 catcagtgagaaaacggctgcatattggtgtcaaagtgtcactgaactaaaggctgactt 1943

Qy.    424 CCCAGACAACGT 435
           ||||||||| |
Db    1944 tccagacaacat 1955


RESULT   5
AAT03143
ID   AAT03143 standard; cDNA; 4369 BP.
XX
AC   AAT03143;
XX
DT   05-JUN-1996  (first entry)
XX
DE   Human dihydropyrimidine dehydrogenase cDNA.
XX
KW   Bovine; liver; human; dihydropyrimidine dehydrogenase; DPD;
KW   pyrimidine catabolism; 5,6-dihydropyrimidine; pyrimidine analogue;
KW   fluoropyrimidine; anticancer drug; 5-fluorouracil; FUra; cancer;
KW   frameshift mutation; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             49..3125
FT                   /*tag= a
FT                   /product= Human lymphocyte DPD
FT   primer_bind     complement (36..60)
FT                   /*tag= b
FT   primer_bind     complement (716..743)
FT                   /*tag= c
FT   primer_bind     1260..1288
FT                   /*tag= d
FT   primer_bind     3117..3139
FT                   /*tag= e
FT   mutation        1000
FT                   /*tag= f
FT                   /note= "Position of single base deletion in DPD
FT                          deficient patients"
XX
PN   WO9528489-A1.
XX
PD   26-OCT-1995.
XX
PF   13-APR-1995;   95WO-US04567.
XX
PR   13-APR-1994;   94US-0227357.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Cheng X,  Diasio RB,  Johnson M,  Lu Z,  Zhang R;
XX
DR   WPI; 1995-373803/48.
DR   P-PSDB; AAR86001.
XX
PT   Novel di:hydro:pyrimidine dehydrogenase gene - used to optimise
PT   5-fluoro:uracil doses given to cancer patients
XX
PS   Disclosure; Fig 7; 207pp; English.
XX
CC   This sequence represents the cDNA sequence encoding human dihydro-
CC   pyrimidine dehydrogenase (DPD).  This sequence was amplified using
CC   the primer sequences given in AAT03144-47.  DPD catalyses the initial
CC   and rate limiting step in pyrimidine catabolism, the reduction of
CC   pyrimidines to 5,6-dihydropyrimidines.  DPD is a complex enzyme
CC   consisting of two identical subunits, containing FMN, FAD and iron-
CC   sulphur centers, and utilising NADPH as a cofactor.  DPD has also been
CC   shown to catalyse the reduction of various pyrimidine analogues
CC   including the fluoropyrimidine anticancer drug 5-fluorouracil (FUra).
CC   Up to 85% of administered FUra may be catabolised by DPD, and it
CC   therefore governs the effectiveness of FUra as an anticancer drug.
CC   DPD genes or fragments of them may be used in the detection of DPD in a
CC   sample, esp. isolated from a cancer patient.  According to the amount
CC   of DPD detected, a therapeutically effective amount of FUra may be
CC   determined and administered.  DPD deficiency, leading to life-
CC   threatening toxicity on exposure to FUra, in a human caused by a
CC   frameshift mutation may be determined by means of a molecular biological
CC   assay to detect the deletion of an A residue at codon 318 within the
CC   DPD-coding region.
XX
SQ   Sequence 4369 BP; 1325 A; 829 C; 932 G; 1283 T; 0 other;


  Query Match             19.7%;  Score 169.6;  DB 16;  Length 4369;
  Best Local Similarity   92.7%;  Pred. No. 5.5e-29;
  Matches  178;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy     244 CTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCG 303
           | | | | ||||     || || ||||||||||||||||||||||||||| |||||||||
Db    1764 caccaaaactttctctcttgataaggacattgtgacaaatgtttcccccagaatcatccg 1823

Qy     304 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1824 gggaaccacctctggccccatgtatggccctggacaaagctcctttctgaatattgagct 1883

Qy     364 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 423
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1884 catcagtgagaaaacggctgcatattggtgtcaaagtgtcactgaactaaaggctgactt 1943

Qy     424 CCCAGACAACGT 435
           ||||||||| |
Db    1944 tccagacaacat 1955


RESULT   6
AAT03132
ID   AAT03132 standard; DNA; 4414 BP.
XX
AC   AAT03132;
```

```
XX
DT   04-JUN-1996  (first entry)
XX
DE   Bovine dihydropyrimidine dehydrogenase gene.
XX
KW   Bovine; liver; human; dihydropyrimidine dehydrogenase; DPD;
KW   pyrimidine catabolism; 5,6-dihydropyrimidine; pyrimidine analogue;
KW   fluoropyrimidine; anticancer drug; 5-fluorouracil; FUra; cancer;
KW   frameshift mutation; ss.
XX
OS   Bos taurus.
XX
FH   Key             Location/Qualifiers
FT   CDS             68..3145
FT                   /*tag= a
FT                   /product= DPD
XX
PN   WO9528489-A1.
XX
PD   26-OCT-1995.
XX
PF   13-APR-1995;   95WO-US04567.
XX
PR   13-APR-1994;   94US-0227357.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Cheng X,  Diasio RB,  Johnson M,  Lu Z,  Zhang R;
XX
DR   WPI; 1995-373803/48.
DR   P-PSDB; AAR86000.
XX
PT   Novel di:hydro:pyrimidine dehydrogenase gene - used to optimise
PT   5-fluoro:uracil doses given to cancer patients
XX
PS   Claim 4; Page 115-32; 207pp; English.
XX
CC   This sequence represents the DNA sequence encoding bovine liver
CC   dihydropyrimidine dehydrogenase (DPD).  DPD catalyses the initial
CC   and rate limiting step in pyrimidine catabolism, the reduction of
CC   pyrimidines to 5,6-dihydropyrimidines.  DPD is a complex enzyme
CC   consisting of two identical subunits, containing FMN, FAD and iron-
CC   sulphur centers, and utilising NADPH as a cofactor.  DPD has also been
CC   shown to catalyse the reduction of various pyrimidine analogues
CC   including the fluoropyrimidine anticancer drug 5-fluorouracil (FUra).
CC   Up to 85% of administered FUra may be catabolised by DPD, and it
CC   therefore governs the effectiveness of FUra as an anticancer drug.
CC   DPD genes or fragments of them may be used in the detection of DPD in a
CC   sample, esp. isolated from a cancer patient.  According to the amount
CC   of DPD detected, a therapeutically effective amount of FUra may be
CC   determined and administered.  DPD deficiency, leading to life-
CC   threatening toxicity on exposure to FUra, in a human caused by a
CC   frameshift mutation may be determined by means of a molecular biological
CC   assay to detect the deletion of an A residue at codon 318 within the
CC   DPD-coding region.
XX
SQ   Sequence 4414 BP; 1315 A; 872 C; 956 G; 1271 T; 0 other;

  Query Match             17.8%;  Score 153.4;  DB 16;  Length 4414;
  Best Local Similarity   88.8%;  Pred. No. 2.3e-25;
  Matches  166;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy     246 CAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGG 305
           | | | ||||     || || |||| || ||||||||||||| |||| ||||||||||
Db    1785 ccaaaactttctctcttgataaggatatagtgacaaatgtttcacccagaatcatccggg 1844

Qy     306 GAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCA 365
           | ||||||||||||||||||||||||||||||||||||||||| || ||||||||||||||
Db    1845 ggaccacctctggccccatgtatggccctggacaaagctctttcctgaatattgagctca 1904

Qy     366 TCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCC 425
           ||||||| |||||||||||||||||||||||||||||||||||||| || ||||| |
Db    1905 tcagtgaaaaaacggctgcatattggtgtcaaagtgtcactgaactaaaagccgactttc 1964

Qy     426 CAGACAA 432
           |||||||
Db    1965 cagacaa 1971


RESULT   7
AAT03142
ID   AAT03142 standard; DNA; 4430 BP.
XX
AC   AAT03142;
XX
DT   04-JUN-1996  (first entry)
XX
DE   Bovine dihydropyrimidine dehydrogenase cDNA.
XX
KW   Bovine; liver; human; dihydropyrimidine dehydrogenase; DPD;
KW   pyrimidine catabolism; 5,6-dihydropyrimidine; pyrimidine analogue;
KW   fluoropyrimidine; anticancer drug; 5-fluorouracil; FUra; cancer;
KW   frameshift mutation; ss.
XX
OS   Bos taurus.
XX
FH   Key             Location/Qualifiers
FT   CDS             67..3144
FT                   /*tag= a
FT                   /product= Bovine DPD
FT   primer_bind     complement (2032..2051)
FT                   /*tag= b
FT                   /note= "Primer A"
FT   primer_bind     2080..2096
FT                   /*tag= c
FT                   /note= "Primer B"
FT   primer_bind     complement (2035..2078)
FT                   /*tag= d
FT                   /note= "Primer C"
FT   primer_bind     complement (85..104)
FT                   /*tag= e
FT                   /note= "Primer D"
FT   primer_bind     2141..2160
FT                   /*tag= f
FT                   /note= "Primer E"
FT   primer_bind     210..229
FT                   /*tag= g
FT                   /note= "Primer F"
XX
PN   WO9528489-A1.
XX
PD   26-OCT-1995.
XX
PF   13-APR-1995;   95WO-US04567.
XX
PR   13-APR-1994;   94US-0227357.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Cheng X,  Diasio RB,  Johnson M,  Lu Z,  Zhang R;
XX
DR   WPI; 1995-373803/48.
DR   P-PSDB; AAR86000.
XX
PT   Novel di:hydro:pyrimidine dehydrogenase gene - used to optimise
PT   5-fluoro:uracil doses given to cancer patients
XX
PS   Example 2; Figure 6; 207pp; English.
XX
CC   This sequence represents the cDNA sequence encoding bovine liver
CC   dihydropyrimidine dehydrogenase (DPD).  This sequence was isolated using
CC   the primers given in AAT03134-41 and AAT14638.  DPD catalyses the
CC   initial and rate limiting step in pyrimidine catabolism, the reduction
CC   of pyrimidines to 5,6-dihydropyrimidines.  DPD is a complex enzyme
CC   consisting of two identical subunits, containing FMN, FAD and iron-
```

CC sulphur centers, and utilising NADPH as a cofactor. DPD has also been
CC shown to catalyse the reduction of various pyrimidine analogues
CC including the fluoropyrimidine anticancer drug 5-fluorouracil (FUra).
CC Up to 85% of administered FUra may be catabolised by DPD, and it
CC therefore governs the effectiveness of FUra as an anticancer drug.
CC DPD genes or fragments of them may be used in the detection of DPD in a
CC sample, esp. isolated from a cancer patient. According to the amount
CC of DPD detected, a therapeutically effective amount of FUra may be
CC determined and administered. DPD deficiency, leading to life-
CC threatening toxicity on exposure to FUra, in a human caused by a
CC frameshift mutation may be determined by means of a molecular biological
CC assay to detect the deletion of an A residue at codon 318 within the
CC DPD-coding region.
XX
SQ Sequence 4430 BP; 1333 A; 871 C; 955 G; 1271 T; 0 other;

Query Match 17.8%; Score 153.4; DB 16; Length 4430;
Best Local Similarity 88.8%; Pred. No. 2.3e-25;
Matches 166; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```
Qy     246 CAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGG 305
           | | | ||||     || || |||| || ||||||||||||| |||| ||||||||||
Db    1784 ccaaaactttctctcttgataaggatatagtgacaaatgtttcacccagaatcatccggg 1843

Qy     306 GAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCA 365
           | ||||||||||||||||||||||||||||||||||||||| || ||||||||||||||||
Db    1844 ggaccacctctggccccatgtatggccctggacaaagctctttcctgaatattgagctca 1903

Qy     366 TCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCC 425
           ||||||| |||||||||||||||||||||||||||||||||||||||| || ||||| |
Db    1904 tcagtgaaaaaacggctgcatattggtgtcaaagtgtcactgaactaaaagccgactttc 1963

Qy     426 CAGACAA 432
           |||||||
Db    1964 cagacaa 1970
```

RESULT 8
AAF58252
ID AAF58252 standard; DNA; 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.

XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 17.6%; Score 151.4; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 4.6e-25;
Matches 8; Conservative 511; Mismatches 272; Indels 0; Gaps 0;

```
Qy       9 AAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGTAATTTATC 68
           :: ::::::::::::: ::::::   : :: :::: ::   ::: :: :: ::::::::
Db       2 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 61

Qy      69 TGGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTTCAATTCTA 128
           :  :  ::: ::::: ::::::::::  ::::::::::  :: :: ::::: :: :::: ::
Db      62 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwww 121

Qy     129 ATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATAAAATATTG 188
           ::::::: :::::::::::::  : :  :::::: : :  :   :  : ::::::::::
Db     122 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 181

Qy     189 TCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGGAGACTCAA 248
           : ::::  :::: :| : :::::::: ::::::  :::: ::: : ::::  : : : ::
Db     182 wwwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 241

Qy     249 TATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAA 308
           ::: :::: : ::: :: :  : ::: : : :::: :::     ::::: ::|     ::
Db     242 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwww 301

Qy     309 CCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCA 368
             :  : :      :: :::     :  : :::  :  ::: : :::::: :  : :: :
Db     302 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 361

Qy     369 GTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAG 428
           : : ::::    :  ::::: : : :::|  | : : :: ::::   : : ::   :
Db     362 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwww 421

Qy     429 ACAACGTAAGTGTGATAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTA 488
           : ::  ::: : : :::::::: ::::: :: : ::::   :::: ::  : ::: ::::
Db     422 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 481

Qy     489 TTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATCTG 548
           :::::: ::  : :: :::   :::::::::::::: : :::::::::: ::  ::: :
Db     482 wwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 541

Qy     549 CCAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAA 608
             : ::  :::  : ::  ::: ::: :::::::: ::: ::::  : ::|:: : ::::
Db     542 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwww 601

Qy     609 AAACCCACACAATGTGAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATC 668
           :::   : : ::: : ::  : : ::::::::   :   :: ::: ::  :: :: :::
Db     602 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 661

Qy     669 TACATAAGCAGAAGGAAGAGAAATGAAATACTCATTTTATTGAGTTGGCCCCCACTGTAT 728
           :: ::::  : ::  :: : :::: ::: : : :::::::: : ::       : : :::
Db     662 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwww 721

Qy     729 GTGGCTGGTATTTATGAAGGTGATGACCCAGGAAGAAATTGTAAACTATAAACCACTCCA 788
             :   :  ::::::: ::  : :: :   :  :: ::::: :::: :::::  : :  :
Db     722 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 781

Qy     789 AATATAAACCC 799
           ::::::  |||
Db     782 wwwwwwgcccc 792
```

RESULT 9

```
AAF58254
ID   AAF58254 standard; DNA; 936 BP.
XX
AC   AAF58254;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1875.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface  -
XX
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;
```

```
  Query Match             17.6%;  Score 151.4;  DB 22;  Length 936;
  Best Local Similarity   1.0%;  Pred. No. 4.6e-25;
  Matches    8;  Conservative  511;  Mismatches  272;  Indels    0;  Gaps    0;

Qy       9 AAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGTAATTTATC 68
           :: :::::::::::: ::::::   : :: ::::  ::   ::: :: :: ::::::::
Db       2 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 61

Qy      69 TGGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTTCAATTCTA 128
           :  :  ::: ::::: :::::::::  :::::::::  :: :: ::::: :: :::: ::
Db      62 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwww 121

Qy     129 ATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATAAAATATTG 188
           ::::::: :::::::::::::  : :  :::::: : :  :   :   : ::::::::::
Db     122 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 181

Qy     189 TCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGGAGACTCAA 248
           : ::::  :::: :| : :::::::: ::::::  :::: ::: : ::::  : : : ::
Db     182 wwwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 241

Qy     249 TATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAA 308
           ::: :::: : ::: :: :  : ::: : : :::: :::     ::::: ::|     ::
Db     242 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwww 301

Qy     309 CCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCA 368
           :  : :      :: :::     :  : :::  :  ::: : :::::: :  : :: :
Db     302 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 361

Qy     369 GTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAG 428
           : : ::::    :  ::::::  : : :::|  | : : :: ::::   : : ::   :
Db     362 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwww 421

Qy     429 ACAACGTAAGTGTGATAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTA 488
           : ::  ::: : : :::::::: ::::: :: : ::::   :::: ::  : ::: ::::
Db     422 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 481

Qy     489 TTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATCTG 548
           :::::: ::  : :: :::   :::::::::::::: : :::::::::: ::  ::: :
Db     482 wwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 541

Qy     549 CCAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAA 608
             : ::  :::  : ::  ::: ::: :::::::: ::: ::::  : ::|:: : ::::
Db     542 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwww 601

Qy     609 AAACCCACACAATGTGAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATC 668
           :::   : : ::: : ::  : : :::::::   :   :: ::: ::  :: :: :::
Db     602 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 661

Qy     669 TACATAAGCAGAAGGAAGAGAAATGAAATACTCATTTTATTGAGTTGGCCCCCACTGTAT 728
           :: ::::  : ::  :: : :::: ::: : : :::::::: : ::        : : :::
Db     662 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwww 721

Qy     729 GTGGCTGGTATTTATGAAGGTGATGACCCAGGAAGAAATTGTAAACTATAAACCACTCCA 788
             :  : ::::::: ::  : :: :   :  :: ::::: :::: ::::::  : :  :
Db     722 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 781

Qy     789 AATATAAACCC 799
           ::::::  |||
Db     782 wwwwwwgcccc 792
```

```
RESULT  10
AAF58257
ID   AAF58257 standard; DNA; 936 BP.
XX
AC   AAF58257;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1954.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface  -
XX
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
```

```
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;


  Query Match             17.6%;  Score 151.4;  DB 22;  Length 936;
  Best Local Similarity   1.0%;  Pred. No. 4.6e-25;
  Matches    8;  Conservative  511;  Mismatches  272;  Indels    0;  Gaps    0;

Qy      9 AAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGTAATTTATC 68
          :: :::::::::::: ::::::   : :: ::::  ::   ::: :: :: ::::::::
Db      2 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 61

Qy     69 TGGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTTCAATTCTA 128
          :  :  ::: ::::: :::::::::  :::::::::  :: :: ::::: :: :::: ::
Db     62 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwww 121

Qy    129 ATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATAAAATATTG 188
          ::::::: :::::::::::::  : :  :::::: : :  :   :  : ::::::::::
Db    122 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 181

Qy    189 TCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGGAGACTCAA 248
          : ::::  :::: :| : :::::::: ::::::  :::: ::: : ::::  : : : ::
Db    182 wwwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 241

Qy    249 TATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAA 308
          ::: :::: : ::: :: :  : ::: : : :::: :::     ::::: ::|     ::
Db    242 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwww 301

Qy    309 CCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCA 368
            : : :      :: :::     :  : :::  :  ::: : :::::: :  : :: :
Db    302 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 361

Qy    369 GTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAG 428
           : : ::::    :  ::::  : : :::|  | : : :: ::::   : : ::   :
Db    362 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwww 421

Qy    429 ACAACGTAAGTGTGATAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTA 488
          : ::  ::: : : :::::::: ::::: :: : ::::   :::: ::  : ::: ::::
Db    422 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 481

Qy    489 TTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATCTG 548
          :::::: ::  : :: :::   :::::::::::::: : :::::::: ::  ::: :
Db    482 wwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 541

Qy    549 CCAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAA 608
            : ::  :::  : ::  ::: ::: :::::::: ::: ::::  : ::|:: : ::::
Db    542 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwww 601

Qy    609 AAACCCACACAATGTGAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATC 668
          :::   : : ::: : ::  : : ::::::::   :   :: ::: ::  :: :: :::
Db    602 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 661

Qy    669 TACATAAGCAGAAGGAAGAGAAATGAAATACTCATTTTATTGAGTTGGCCCCCACTGTAT 728
          :: ::::  : ::  :: : :::: ::: : : :::::::: : ::        : : :::
Db    662 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwww 721

Qy    729 GTGGCTGGTATTTATGAAGGTGATGACCCAGGAAGAAATTGTAAACTATAAACCACTCCA 788
           :   :  ::::::: ::  : :: :   :  :: ::::: :::: ::::::  : :  :
Db    722 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 781

Qy    789 AATATAAACCC 799
          ::::::  |||
Db    782 wwwwwwgcccc 792


RESULT  11
AAF58259
ID   AAF58259 standard; DNA; 936 BP.
XX
AC   AAF58259;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D2004.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface  -
XX
PS   Example 6; Page 128; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;


  Query Match             17.6%;  Score 151.4;  DB 22;  Length 936;
  Best Local Similarity   1.0%;  Pred. No. 4.6e-25;
  Matches    8;  Conservative  511;  Mismatches  272;  Indels    0;  Gaps    0;

Qy      9 AAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGTAATTTATC 68
          :: :::::::::::: ::::::   : :: ::::  ::   ::: :: :: ::::::::
Db      2 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 61

Qy     69 TGGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTTCAATTCTA 128
          :  :  ::: ::::: :::::::::  :::::::::  :: :: ::::: :: :::: ::
Db     62 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwww 121

Qy    129 ATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATAAAATATTG 188
          ::::::: :::::::::::::  : :  :::::: : :  :   :  : ::::::::::
Db    122 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 181

Qy    189 TCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGGAGACTCAA 248
          : ::::  :::: :| : :::::::: ::::::  :::: ::: : ::::  : : : ::
Db    182 wwwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 241

Qy    249 TATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAA 308
          ::: :::: : ::: :: :  : ::: : : :::: :::     ::::: ::|     ::
Db    242 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwww 301

Qy    309 CCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCA 368
            : : :      :: :::     :  : :::  :  ::: : :::::: :  : :: :
Db    302 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 361

Qy    369 GTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAG 428
           : : ::::    :  ::::  : : :::|  | : : :: ::::   : : ::   :
Db    362 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwww 421

Qy    429 ACAACGTAAGTGTGATAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTA 488
```

```
                : ::  ::: : : :::::::: ::::: :: : ::::   :::: ::  : ::: ::::
Db    422 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 481

Qy    489 TTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATCTG 548
          :::::: ::  : :: :::   ::::::::::::: : :::::::::: ::   ::: :
Db    482 wwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 541

Qy    549 CCAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAA 608
             : ::  :::  : ::  ::: ::: :::::::: ::: ::::  : ::|:: : ::::
Db    542 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwww 601

Qy    609 AAACCCACACAATGTGAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATC 668
          :::    : : ::: : ::  : : ::::::::   :    :: ::: ::  :: :: :::
Db    602 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 661

Qy    669 TACATAAGCAGAAGGAAGAGAAATGAAATACTCATTTTATTGAGTTGGCCCCCACTGTAT 728
          :: ::::   : ::   :: : :::: ::: : : :::::::: : ::          : : :::
Db    662 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwww 721

Qy    729 GTGGCTGGTATTTATGAAGGTGATGACCCAGGAAGAAATTGTAAACTATAAACCACTCCA 788
            :    :  ::::::: ::   : :: :    :   :: ::::: :::: ::::::  : :  :
Db    722 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 781

Qy    789 AATATAAACCC 799
          ::::::  |||
Db    782 wwwwwwgcccc 792
```

RESULT 12
AAF58262

```
ID   AAF58262 standard; DNA; 936 BP.
XX
AC   AAF58262;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D2007.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface  -
XX
PS   Example 6; Page 128; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;
```

```
  Query Match             17.6%;  Score 151.4;  DB 22;  Length 936;
  Best Local Similarity    1.0%;  Pred. No. 4.6e-25;
  Matches    8;  Conservative  511;  Mismatches  272;  Indels    0;  Gaps    0;

Qy      9 AAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGTAATTTATC 68
          :: :::::::::::: ::::::   : :: ::::  ::    ::: :: :: ::::::::
Db      2 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 61

Qy     69 TGGAGCTAACAAATACTTTATTTACCTTTTTATTTGCAAGTAGTTTATGTTCAATTCTA 128
           :   :  ::: ::::: :::::::::  :::::::::  :: :: ::::: :: :::: ::
Db     62 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwww 121

Qy    129 ATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATAAAATATTG 188
          :::::: :::::::::::: : :   :::::: : :  :    :  : :::::::::
Db    122 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 181

Qy    189 TCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGGAGACTCAA 248
           : ::::  :::: :| : :::::::: ::::::  :::: ::: : ::::  : : : ::
Db    182 wwwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 241

Qy    249 TATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAA 308
          ::: :::: : ::: :: :  : ::: : : :::: :::     ::::: ::|      ::
Db    242 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwww 301

Qy    309 CCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCA 368
             :  : :      :: :::     :  : :::  :  ::: : :::::: :  : :: :
Db    302 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 361

Qy    369 GTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAG 428
            : : ::::    :  :::::  : : :::|   | : : :: ::::   : : ::   :
Db    362 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwww 421

Qy    429 ACAACGTAAGTGTGATAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTA 488
           : ::  ::: : : :::::::: ::::: :: : ::::    :::: ::  : ::: ::::
Db    422 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 481

Qy    489 TTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATCTG 548
          :::::: ::  : :: :::   ::::::::::::: : :::::::::: ::   ::: :
Db    482 wwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 541

Qy    549 CCAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAA 608
             : ::  :::  : ::  ::: ::: :::::::: ::: ::::  : ::|:: : ::::
Db    542 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwww 601

Qy    609 AAACCCACACAATGTGAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATC 668
          :::    : : ::: : ::  : : ::::::::   :    :: ::: ::  :: :: :::
Db    602 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 661

Qy    669 TACATAAGCAGAAGGAAGAGAAATGAAATACTCATTTTATTGAGTTGGCCCCCACTGTAT 728
          :: ::::   : ::   :: : :::: ::: : : :::::::: : ::          : : :::
Db    662 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwww 721

Qy    729 GTGGCTGGTATTTATGAAGGTGATGACCCAGGAAGAAATTGTAAACTATAAACCACTCCA 788
            :    :  ::::::: ::   : :: :    :   :: ::::: :::: ::::::  : :  :
Db    722 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 781

Qy    789 AATATAAACCC 799
          ::::::  |||
Db    782 wwwwwwgcccc 792
```

RESULT 13
AAF58255

```
ID   AAF58255 standard; DNA; 938 BP.
XX
AC   AAF58255;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1876.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
```

```
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface  -
XX
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

  Query Match             17.6%;  Score 151.4;  DB 22;  Length 938;
  Best Local Similarity   1.0%;  Pred. No. 4.6e-25;
  Matches    8;  Conservative  511;  Mismatches  272;  Indels    0;  Gaps    0;

Qy      9 AAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGTAATTTATC 68
            :: ::::::::::: ::::::    : :: ::::  ::    ::: :: :: ::::::::
Db      2 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 61

Qy     69 TGGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTTCAATTCTA 128
            :   :  ::: ::::: :::::::::  :::::::::  :: :: ::::: :: :::: ::
Db     62 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwww 121

Qy    129 ATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATAAAATATTG 188
            ::::::: :::::::::::::   : :  :::::: : :   :    :   : :::::::::
Db    122 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 181

Qy    189 TCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGGAGACTCAA 248
            : ::::   :::: :| : :::::::: ::::::  :::: ::: : ::::   : : : ::
Db    182 wwwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 241

Qy    249 TATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAA 308
            ::: :::: : ::: :: :   : ::: : : :::: :::        ::::: ::|     ::
Db    242 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwww 301

Qy    309 CCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCA 368
               :  : :       :: :::       :   : :::  :  ::: : :::::: :  : :: :
Db    302 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 361

Qy    369 GTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAG 428
             : : ::::     :  :::::  : : :::|  | : : :: ::::   : : ::    :
Db    362 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwww 421

Qy    429 ACAACGTAAGTGTGATAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTA 488
            : ::  ::: : : :::::::: ::::: :: : ::::    :::: ::   : ::: ::::
Db    422 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 481

Qy    489 TTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATCTG 548
            :::::: ::   : :: :::    :::::::::::::: : :::::::::: ::   ::: :

Db    482 wwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 541

Qy    549 CCAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAA 608
              : ::  :::   : ::   ::: ::: :::::::: ::: ::::   : ::|:: : ::::
Db    542 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwww 601

Qy    609 AAACCCACACAATGTGAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATC 668
            :::    : : ::: : ::   : : ::::::::    :     :: ::: ::   :: :: :::
Db    602 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 661

Qy    669 TACATAAGCAGAAGGAAGAGAAATGAAATACTCATTTTATTGAGTTGGCCCCCACTGTAT 728
            :: ::::   : ::   :: : :::: ::: : : :::::::: : ::           : : :::
Db    662 wwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 721

Qy    729 GTGGCTGGTATTTATGAAGGTGATGACCCAGGAAGAAATTGTAAACTATAAACCACTCCA 788
              :    :  ::::::: ::    : :: :     :   :: ::::: :::: ::::::   : :  :
Db    722 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 781

Qy    789 AATATAAACCC 799
            ::::::  |||
Db    782 wwwwwwgcccc 792


RESULT 14
AAT14078
ID   AAT14078 standard; cDNA; 4447 BP.
XX
AC   AAT14078;
XX
DT   11-JUN-1996  (first entry)
XX
DE   Pig dihydropyrimidine dehydrogenase cDNA.
XX
KW   Dihydropyrimidine dehydrogenase; DPD; DPYD gene; probe; deficiency;
KW   5-fluorouracil; cytostatic; cancer; selectable marker; ss.
XX
OS   Sus scrofa.
XX
FH   Key             Location/Qualifiers
FT   CDS             88..3165
FT                   /*tag=  a
XX
PN   WO9608568-A2.
XX
PD   21-MAR-1996.
XX
PF   07-SEP-1995;   95WO-US12016.
XX
PR   12-SEP-1994;   94US-0304309.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Fernandez-Salguero P,  Gonzalez FJ;
XX
DR   WPI; 1996-179943/18.
DR   P-PSDB; AAR91421.
XX
PT   Pig and human di:hydro:pyrimidine dehydrogenase (DPD) genes and
PT   probes  - useful for detection of DPD deficiencies and
PT   identification of humans at risk of toxic reaction to
PT   5-fluoro:uracil anti-cancer treatment
XX
PS   Claim 1; Page 52-56; 78pp; English.
XX
CC   Pig dihydropyrmidine dehydrogenase (DPD) cDNA (AAT14078) was
CC   isolated from a pig liver cDNA library by screening with a
CC   polyclonal antibody raised against pig DPD, and rescreening
CC   using an isolated partial cDNA.  The DPD cDNA can be inserted
CC   into a vector and used for prodn. of active DPD (AAR91421) in
CC   transformed host cells, e.g. Escherichia coli, or used as a
CC   selectable marker functional in prokaryotic and eukaryotic cells.
CC   It can also be utilised as a probe to detect DPD deficiency in
CC   an individual, i.e. to identify risk of toxic reaction to
```

```
CC   4-fluorouracil.
XX
SQ   Sequence 4447 BP; 1285 A; 928 C; 1000 G; 1234 T; 0 other;

  Query Match           17.4%;  Score 150.2;  DB 17;  Length 4447;
  Best Local Similarity  87.7%;  Pred. No. 1.2e-24;
  Matches  164;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy     246 CAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGG 305
           | | | ||||     || || ||||||| |||||||||| || |||| |||| ||||||
Db    1805 ccaaaactttctctcttgataaggacatagtgacaaatgtctcacccagaatcgtccggg 1864

Qy     306 GAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCA 365
           | || ||||||||||||||||| |||||||||||||||||||| ||||||||||||||||
Db    1865 ggactacctctggccccatgtacggccctggacaaagctccttcctgaatattgagctca 1924

Qy     366 TCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCC 425
           ||||||| ||||| |||||||||||||||||||||||||||||||||| |||||||| |
Db    1925 tcagtgaaaaaacagctgcatattggtgtcaaagtgtcactgaactaaaagctgactttc 1984

Qy     426 CAGACAA 432
           |||||||
Db    1985 cagacaa 1991


RESULT  15
AAF58252/c
ID   AAF58252 standard; DNA; 936 BP.
XX
AC   AAF58252;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1835.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface -
XX
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

  Query Match           17.1%;  Score 147.4;  DB 22;  Length 936;
  Best Local Similarity   0.5%;  Pred. No. 3.6e-24;
  Matches    4;  Conservative  510;  Mismatches  271;  Indels    0;  Gaps    0;

Qy      12 ATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGTAATTTATCTGG 71
           :::::::::::: ::::::   : :: :::  ::   ::: :: :: :::::::: :
Db     786 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 727

Qy      72 AGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTTCAATTCTAATT 131
           :  ::: ::::: :::::::::  :::::::::  :  :: ::::: :: :::: :::::
Db     726 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWW 667

Qy     132 TAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATAAAATATTGTCA 191
           :::: :::::::::::::  : :  :::::: : :  :   :  : :::::::::: : :
Db     666 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 607

Qy     192 TATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGGAGACTCAATAT 251
           :::  :::: :  : :::::::: ::::::  :::: ::: : :::  : : : :::::
Db     606 WWWWWWWWWWWWWWCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 547

Qy     252 CTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAACCA 311
           :::: : ::: :: :  : ::: : : :::: :::     ::::: ::   |  ::  :
Db     546 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWW 487

Qy     312 CCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCAGTG 371
             : :      :: :::    :  : :::  :  ::: : ::::::  :  : :: : :
Db     486 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 427

Qy     372 AGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAGACA 431
           : ::::    :  ::::: : : ::: : | |   :: ::::   : : ::   : : :
Db     426 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWTAAGCWWWWWWWWWWWWWWWWWWWWWWW 367

Qy     432 ACGTAAGTGTGATAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTATTT 491
           :  ::: : : :::::::: ::::: :: : ::::   :::: ::  : ::: :::::::
Db     366 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 307

Qy     492 AAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATCTGCCA 551
           ::: ::  :::: :::   :::::::::::::: : :::::::::: ::  ::: :   :
Db     306 WWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 247

Qy     552 GTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAAAAA 611
           ::  :::  : ::  ::: ::: :::::::: ::: :::  : :: ::|: :::::::
Db     246 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWCWWWWWWWW 187

Qy     612 CCCACACAATGTGAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATCTAC 671
             : : ::: : ::  : : :::::::   :   :: ::: ::  :: :: ::: ::
Db     186 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 127

Qy     672 ATAAGCAGAAGGAAGAGAAATGAAATACTCATTTTATTGAGTTGGCCCCCACTGTATGTG 731
           ::::  : ::  :: : :::: :::::   :::::::: : ::        : : ::: :
Db     126 WWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWW 67

Qy     732 GCTGGTATTTATGAAGGTGATGACCCAGGAAGAAATTGTAAACTATAAACCACTCCAAAT 791
           :  :::::: ::  : :: :   :  :: ::::: :::: ::::::  : :  ::::
Db      66 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 7

Qy     792 ATAAA 796
           :::::
Db       6 WWWWW 2


Search completed: November 20, 2001, 15:59:11
Job time: 5375 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2001, 14:25:51 ; Search time 1355.35 Seconds
(without alignments)
6005.022 Million cell updates/sec

Title: US-09-308-080-1
Perfect score: 861
Sequence: 1 TGTTAATGAAGATAAATATT.....AGTGGGAAATAATTTATTAA 861

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*
190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	169.6	19.7	938	172	BG024298	BG024298 602274460
c 2	92.4	10.7	853	220	CNS02EEI	AL193635 Tetraodon
3	86.2	10.0	621	110	AW018806	AW018806 fd60d03.y
4	70	8.1	1207	219	CNS00175	AL058798 Drosophil
5	62.6	7.3	562	165	BE288474	BE288474 601095732
6	58	6.7	1101	219	CNS0039G	AL063921 Drosophil
7	55	6.4	500	107	AU086287	AU086287 AU086287
8	53.4	6.2	483	151	BF653737	BF653737 277669 MA
c 9	52.6	6.1	1101	219	CNS00EVL	AL069706 Drosophil
10	52.6	6.1	1101	219	CNS00Z0C	AL097062 Drosophil
c 11	52.4	6.1	1025	219	CNS014J2	AL104216 Drosophil
12	52	6.0	912	245	AZ527678	AZ527678 ENTBN23TR
13	52	6.0	1101	219	CNS00EO7	AL069440 Drosophil
c 14	51.8	6.0	853	245	AZ548198	AZ548198 ENTFW05TF
c 15	51.8	6.0	868	248	AZ686179	AZ686179 ENTJI41TR
c 16	51.8	6.0	881	247	AZ681140	AZ681140 ENTMC34TR
17	51.8	6.0	899	245	AZ542627	AZ542627 ENTCX21TR
18	51.4	6.0	898	248	AZ686287	AZ686287 ENTKG36TR
19	51.2	5.9	987	219	CNS014PQ	AL104456 Drosophil
20	50	5.8	632	107	AU060843	AU060843 AU060843
c 21	49.8	5.8	947	247	AZ682793	AZ682793 ENTKY77TF
22	49.8	5.8	1101	219	CNS016LI	AL106896 Drosophil
23	49.6	5.8	867	219	CNS00CX5	AL060052 Drosophil
c 24	49.4	5.7	872	245	AZ535541	AZ535541 ENTBP32TR
c 25	49.4	5.7	1101	219	CNS0039G	AL063921 Drosophil
c 26	49	5.7	896	245	AZ534072	AZ534072 ENTBR33TR
c 27	48.6	5.6	872	245	AZ527849	AZ527849 ENTCF51TR
28	48.6	5.6	1101	219	CNS017V6	AL108540 Drosophil
29	48.4	5.6	329	105	AL513719	AL513719 AL513719
30	48.4	5.6	911	248	AZ686515	AZ686515 ENTIU42TR
31	48.2	5.6	898	247	AZ674122	AZ674122 ENTIB47TF
c 32	48.2	5.6	987	219	CNS014PQ	AL104456 Drosophil
33	48.2	5.6	1007	219	CNS01460	AL103746 Drosophil
c 34	48	5.6	892	245	AZ549872	AZ549872 ENTFG30TR
c 35	47.8	5.6	1101	219	CNS003DQ	AL064580 Drosophil
c 36	47.4	5.5	872	248	AZ683520	AZ683520 ENTIO43TR
c 37	47.4	5.5	878	219	CNS0187R	AL108993 Drosophil
c 38	47.2	5.5	816	220	CNS02KJV	AL201604 Tetraodon
c 39	47.2	5.5	909	248	AZ690408	AZ690408 ENTLW08TR
40	47.2	5.5	1032	220	CNS02OIP	AL206746 Tetraodon
c 41	47.2	5.5	1048	248	AZ683764	AZ683764 ENTIL96TR
42	46.8	5.4	959	219	CNS00655	AL062806 Drosophil
43	46.8	5.4	1021	219	CNS014DY	AL104032 Drosophil
c 44	46.8	5.4	1092	220	CNS020K7	AL175696 Tetraodon
c 45	46.8	5.4	1201	219	CNS016AD	AL106495 Drosophil

ALIGNMENTS

```
RESULT   1
BG024298
LOCUS       BG024298      938 bp    mRNA            EST       24-JAN-2001
DEFINITION  602274460F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4362302 5',
            mRNA sequence.
ACCESSION   BG024298
VERSION     BG024298.1  GI:12409730
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 938)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis Staudt, M.D., Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10006  row: h  column: 15
            High quality sequence stop: 714.
FEATURES             Location/Qualifiers
     source          1. .938
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4362302"
                     /clone_lib="NIH_MGC_85"
                     /tissue_type="lymphoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.867 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      222 a    234 c    253 g    229 t
ORIGIN


  Query Match             19.7%;  Score 169.6;  DB 172;  Length 938;
  Best Local Similarity   92.7%;  Pred. No. 7.9e-29;
  Matches  178;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy     244 CTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCG 303
           | | | | ||||     || || |||||||||||||||||||||||||||| |||||||||
Db     260 CACCAAAACTTTCTCTCTTGATAAGGACATTGTGACAAATGTTTCCCCCAGAATCATCCG 319

Qy     304 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     320 GGGAACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 379

Qy     364 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 423
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     380 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 439

Qy     424 CCCAGACAACGT 435
           ||||||||| |
Db     440 TCCAGACAACAT 451


RESULT   2
CNS02EEI/c
LOCUS       CNS02EEI      853 bp    DNA             GSS       13-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            262A08 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL193635
VERSION     AL193635.1  GI:7831741
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 853)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 853)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
```

```
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 853)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .853
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="262A08"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG262BA04LP1-end : T7"
BASE COUNT      252 a    189 c    136 g    269 t      7 others
ORIGIN
```

Query Match 10.7%; Score 92.4; DB 220; Length 853;
Best Local Similarity 68.5%; Pred. No. 3.8e-11;
Matches 126; Conservative 1; Mismatches 57; Indels 0; Gaps 0;

```
Qy     258 TCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAACCACCTCTG 317
           | ||||   ||||| | || |||||||| || ||    ||  | || || ||||| || |
Db     409 TGTTTCTGCAGGACTTGGTAACAAATGTGTCTCCTCGTATTGTGCGAGGCACCACTTCAG 350

Qy     318 GCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCAGTGAGAAAA 377
           ||    |||||||||| || ||  | || || || || |||||||| || || ||||| |
Db     349 GCAATGTGTATGGCCCAGGTCAGGGGTCATTCCTCAACATTGAGCTTATTAGCGAGAAGA 290

Qy     378 CGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAGACAACGTAA 437
           |||| || ||||||||||||   ||  | |:||| |||     ||| ||  | || ||||
Db     289 CGGCAGCTTATTGGTGTCAATCAGTGGCCGMACTTAAGAGGAGCTTTCCGAATAATGTAA 230

Qy     438 GTGT 441
           || |
Db     229 GTTT 226
```

```
RESULT   3
AW018806
LOCUS       AW018806      621 bp    mRNA            EST       13-SEP-1999
DEFINITION  fd60d03.y1 Zebrafish WashU MPIMG EST Danio rerio cDNA 5' similar to
            SW:DPYD_HUMAN Q12882 DIHYDROPYRIMIDINE DEHYDROGENASE [NADP+]
            PRECURSOR ;, mRNA sequence.
ACCESSION   AW018806
VERSION     AW018806.1  GI:5872336
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 621)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 491.
FEATURES             Location/Qualifiers
     source          1. .621
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone_lib="Zebrafish WashU MPIMG EST"
                     /sex="mixed"
                     /tissue_type="26 somite embryos, adult livers, shield
                     stage embryos"
                     /lab_host="XL1-blue MRF"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
                     strand cDNA was primed with a Not I - oligo(dT)15 primer
                     [5'pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTT3'];
                     double-stranded cDNA was ligated to Sal I adaptors (BRL),
                     digested with Not I and cloned into the Not I and Sal I
                     sites of the pSPORT1 vector (BRL). Library was constructed
                     by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
                     Institut fuer Molekulare Genetik,Berlin). cDNAs for EST
                     analysis were selected following oligonucleotide
                     hybridization fingerprinting of arrayed clones from
                     zebrafish late somitogenesis (26 ss), adult liver or
                     embryonic shield stage (5.6 h) libraries.  Fingerprint
                     data were used to computationally cluster cDNAs, and a
                     single cDNA from each cluster was chosen for sequencing.
                     In some cases multiple members of the same cluster were
                     sequenced to assess clustering parameters or single clones
                     were sequenced additional times to assess quality
                     control."
BASE COUNT      151 a    169 c    160 g    141 t
ORIGIN
```

Query Match 10.0%; Score 86.2; DB 110; Length 621;
Best Local Similarity 70.6%; Pred. No. 1e-09;
Matches 115; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

```
Qy     275 GTGACAAATGTTTCCCCCATAATCATCCGGGGAACCACCTCTGGCCCCATGTATGGCCCT 334
           ||||| || || || ||   |||| | || || || ||||| || | ||| |  || |||
Db       1 GTGACCAACGTGTCACCGCGAATCGTGCGTGGGACAACCTCAGGTCACATATTCGGTCCT 60

Qy     335 GGACAAAGCTCCTTTCTGAATATTGAGCTCATCAGTGAGAAAACGGCTGCATATTGGTGT 394
           || ||  |||| || || || || |||||||| ||||| ||||| || || || ||||| 
Db      61 GGTCAGGGCTCATTCCTCAACATCGAGCTCATTAGTGAAAAAACTGCCGCTTACTGGTGC 120

Qy     395 CAAAGTGTCACTGAACTAAAGGCTGACTTCCCAGACAACGTAA 437
            | || ||  ||||||| || || ||||||||  | ||  | |
Db     121 AAGAGCGTGGCTGAACTCAAAGCCGACTTCCCCAAAAATATTA 163
```

```
RESULT   4
CNS00175
LOCUS       CNS00175     1207 bp    DNA             GSS       04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR03G12 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL058798
VERSION     AL058798.1  GI:4939103
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1207)
```

```
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .1207
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone_lib="RPCI-98"
                      /clone="BACR03G12"
                      /note="end : TET3"
BASE COUNT      250 a    298 c    328 g    233 t     98 others
ORIGIN

  Query Match             8.1%;  Score 70;  DB 219;  Length 1207;
  Best Local Similarity   63.9%;  Pred. No. 5.3e-06;
  Matches  106;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy    267 AGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAACCACCTCTGGCCCCATGT 326
          ||||  | || || ||||| || ||    ||| || |||| ||||| || |||  || ||
Db    399 AGGATCTGGTCACGAATGTCTCGCCGCGCATCGTCAGGGGCACCACGTCGGGCTACAAGT 458

Qy    327 ATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCAGTGAGAAAACGGCTGCAT 386
          |||| ||    ||  |||  || ||||| || || || |||   |||||   ||| |  |
Db    459 ATGGACCGCAGCAGGGCTGTTTCCTGAACATCGAACTAATCTCGGAGAAGCGGGCCGAGT 518

Qy    387 ATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAGACAA 432
          | |||    ||    |    ||||| |||  ||||||||| || ||
Db    519 ACTGGCTGAAATCGATTGGAGAACTGAAGCGTGACTTCCCCGAGAA 564


RESULT   5
BE288474
LOCUS       BE288474      562 bp    mRNA            EST       26-OCT-2000
DEFINITION  601095732F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3490423 5',
            mRNA sequence.
ACCESSION   BE288474
VERSION     BE288474.1  GI:9168194
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 562)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8532  row: p  column: 08
            High quality sequence stop: 562.
FEATURES              Location/Qualifiers
     source           1. .562
                      /organism="Mus musculus"
                      /strain="C57/B6"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:3490423"
                      /clone_lib="NCI_CGAP_Mam5"
                      /tissue_type="tumor, gross tissue"
                      /dev_stage="7 months"
                      /lab_host="DH10B"
                      /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                      Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                      Library constructed by Life Technologies. Investigators
                      providing samples: Lothar Hennighausen/Robin Humphreys,
                      NIH"
BASE COUNT      137 a    124 c    166 g    135 t
ORIGIN


  Query Match             7.3%;  Score 62.6;  DB 165;  Length 562;
  Best Local Similarity   94.2%;  Pred. No. 0.00025;
  Matches   65;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy    364 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTT 423
          |||||||||||||||| |||||||||||||||| |||||||| ||||||||||||||||
Db      1 CATCAGTGAGAAAACAGCTGCATATTGGTGTCACAGTGTCACCGAACTAAAGGCTGACTT 60

Qy    424 CCCAGACAA 432
          ||| |||||
Db     61 CCCGGACAA 69


RESULT   6
CNS0039G
LOCUS       CNS0039G     1101 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .1101
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
```

```
                     /clone_lib="RPCI-98"
                     /clone="BACR08K10"
                     /note="end : TET3"
BASE COUNT       201 a      64 c     131 g     202 t     503 others
ORIGIN
```

```
  Query Match             6.7%;  Score 58;  DB 219;  Length 1101;
  Best Local Similarity   20.4%;  Pred. No. 0.003;
  Matches  116;  Conservative  216;  Mismatches  232;  Indels    6;  Gaps    1;

Qy      70 GGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTTCAATTCTAA 129
           || |:  | |:|:: ::| ||||   :   :|     |  ::  : :     | |:  ||
Db     391 GGGGMATATAWAWWWWWTTTTTTTTAWAAAWAAAATAATTWWAAWAWAAAAAATTWWAAA 450

Qy     130 TTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATAAAATATTGT 189
           : ||   :||::   |:|:       ||| || :             |:| |::|| :
Db     451 AWAAAAWAWTAWWTTTAWTWAAAAAAAAAAAATTWTTTTTTTWTTTAWTTWATAWWTTWW 510

Qy     190 CATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGGAGACTCAAT 249
             ||   |||  |  | |:|| :|| : || |  |::| :|| | :||:  | |: :||:
Db     511 WTTAAAWAAAAAAAAAAAAWAAAWAAAWATAAATWTWWTWWTTYTTWAAWATAAAMCMAAW 570

Qy     250 ATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGGAAC 309
           ::|:| :|:::: :      :::   :|:|: |::: ::   ::: :| :    |:
Db     571 YYHTYTTYTYHYYTYWTYTMTWHYHTMYTHAWAHTTWYHWYHTYAMWHWMTWHWTMWAWW 630

Qy     310 CACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCAG 369
           ::  :   ::::  : :  :: :  : :||  :  :|: | | :: | : ::::::||
Db     631 HWTTYTAAYYYYYTCMYYYHYMHWHHAHAHAAAWWTTHTWWTHAYHWATYHYYYYMYCAM 690

Qy     370 TGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCCAGA 429
             :  :  :  :: ::: |   | |:::  : ::::: :|::::       || | :: :
Db     691 MCMCTHTCHHCYYYYHHYTAHHTHTHHWYAHYYMWYYMWAYYWM------YCTACTYHYH 744

Qy     430 CAACGTAAGTGTGATAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTAT 489
           :::: ::| : |   :|:|::::::|::|| | :| :   :: : ||    : : :::
Db     745 HHHHYHWAYHTTWYAWAHAMWMWHHAHYAAAAAWAAWATTHHYHHTTHYMHHTYMYHYYM 804

Qy     490 TTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATCTGC 549
           :|   : |:: :::::|| |:|:|::: :|: :: : |::::|||:: ::: : :|:
Db     805 YTCCYMCTYHCWHYYHTAYTCWTWTHHWMWTWTHWYHHTWWHHTTTHWAWWHTHTWCWWW 864

Qy     550 CAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAAA 609
           : ||  : | |:    :::: ::: :::::| : ::: :: : :::::   ::::|: :
Db     865 WHATTWTWATHCWACMTMHWHHWMHMHHHHMACHAHHHTHMCMCHHHMHCTCHHHHTMYH 924

Qy     610 AACCCACACAATGTGAAGCTCTGTTATAAA 639
           : |: : ::::: : ::  |   |  |:::
Db     925 MTCHWMWHMHWHWHHWWMATWMTTMTTMMM 954
```

```
RESULT   7
AU086287
LOCUS       AU086287      500 bp    mRNA            EST       27-JAN-2001
DEFINITION  AU086287 Sugano Malaria cDNA library Plasmodium falciparum cDNA
            clone XPFn2249, mRNA sequence.
ACCESSION   AU086287
VERSION     AU086287.1  GI:12388428
KEYWORDS    EST.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.
  TITLE     FULL-malaria: a database for a full-length enriched cDNA library
            from human malaria parasite, Plasmodium falciparum
  JOURNAL   Nucleic Acids Res. 29 (1), 70-71 (2001)
  MEDLINE   20574754
COMMENT     Contact: Junichi Watanabe
            Institute of Medical Science
            The University of Tokyo, Department of Parasitology
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Tel: 81-3-5449-5378
            Fax: 81-3-5449-5410
            Email: jwatanab@manage.ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Plasmodium falciparum"
                     /strain="3D7"
                     /db_xref="taxon:5833"
                     /clone="XPFn2249"
                     /clone_lib="Sugano Malaria cDNA library"
                     /dev_stage="erythrocytic stage"
BASE COUNT       214 a      25 c      24 g     231 t       6 others
ORIGIN
```

```
  Query Match             6.4%;  Score 55;  DB 107;  Length 500;
  Best Local Similarity   49.3%;  Pred. No. 0.014;
  Matches  136;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy       7 TGAAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGTAATTTA 66
           | || || | | |||||  |||||       | |||         | || | |  || ||
Db     202 TTAATATTATTTTTTTTCATTTTTGTAATATATAATATATATTTGACTACATTTATTTAT 261

Qy      67 TCTGGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTTCAATTC 126
           | |    || ||| ||| ||||||||||   ||||    | |  |  | || ||||||
Db     262 TATATTTCTGACATATAATTTATTTTATTATTTTTAAAGAATTTNNTATAAGTTCAAAAT 321

Qy     127 TAATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATAAAATAT 186
           |||  |||  ||||  |  ||   | |  | | |        |   |  | ||| || ||
Db     322 AAATAAAATCCATATATATTATATATATAAATAAAACATTTTATAAATATTATATAAAAT 381

Qy     187 TGTCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGGAGACTC 246
           | |||  | |    || ||||  |   |||||  |  |  | |  | |         |
Db     382 CTCCTTATAAACAAATACAAATAAATATATTTATATATAATTATAGATATATTTTTGTGC 441

Qy     247 AATATCTTTACTCTTTCATGAGGACATTGTGACAAA 282
           | | | |||  | |||  |        || | | |||
Db     442 ATTTTTTTTNGTTTTTANTTTTTNATTTATAATAAA 477
```

```
RESULT   8
BF653737
LOCUS       BF653737      483 bp    mRNA            EST       20-DEC-2000
DEFINITION  277669 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF653737
VERSION     BF653737.1  GI:11918869
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 483)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
```

```
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 70  row: B  column: 4
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES              Location/Qualifiers
     source           1. .483
                      /organism="Bos taurus"
                      /db_xref="taxon:9913"
                      /clone_lib="MARC 3BOV"
                      /tissue_type="pooled"
                      /lab_host="DH10B"
                      /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                      Library made from pooled tissue from marrow, alveolar
                      macrophage, ovary, fetal semitendonosus muscle, and fetal
                      longissimus muscle."
BASE COUNT      128 a    105 c    138 g    112 t
ORIGIN


  Query Match             6.2%;  Score 53.4;  DB 151;  Length 483;
  Best Local Similarity   90.7%;  Pred. No. 0.032;
  Matches   68;  Conservative    0;  Mismatches    6;  Indels    1;  Gaps    1;

Qy     359 GAGCTCATCAGTG-AGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGC 417
           | |||||||||||| | |||||||||||| |||||||||||||||||||||||||||| ||
Db       1 GCGCTCATCAGTGAAAAAAACGGCTGCGTATTGGTGTCAAAGTGTCACTGAACTAAAAGC 60

Qy     418 TGACTTCCCAGACAA 432
           ||||| ||||||||
Db      61 CGACTTTCCAGACAA 75


RESULT   9
CNS00EVL/c
LOCUS       CNS00EVL     1101 bp    DNA             GSS       04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .1101
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone_lib="RPCI-98"
                      /clone="BACR29B23"
                      /note="end : T7"
BASE COUNT      419 a     91 c     60 g    299 t    232 others
ORIGIN


  Query Match             6.1%;  Score 52.6;  DB 219;  Length 1101;
  Best Local Similarity   35.7%;  Pred. No. 0.051;
  Matches   99;  Conservative   51;  Mismatches  127;  Indels    0;  Gaps    0;

Qy       1 TGTTAATGAAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGT 60
           | |: || |  :| :|||:||||| ||||||   :  | |:||  ||         |: :| |
Db     723 TATWTATWATTWTTWATWTTTTTTTTTTTTTTTTWTATTTWAAATTATTTAATTAWWWAAT 664

Qy      61 AATTTATCTGGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTT 120
           :|:|||| |     : |   | | | |     |  || |:|      | |  |||| ||
Db     663 WAWTTATTTAATAAWTAATTAAATTATTAAAAATTTTATWAAAAAATATTTTTTTATATT 604

Qy     121 CAATTCTAATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATA 180
           | :    | ||:  ||   ||::  :|    |   |||:| | :      |    :|
Db     603 AATWAWATTTATAWATTAATATAWWTTWTAAATTWTTAATWTATTWWTTTAAATTATWTT 544

Qy     181 AAATATTGTCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGG 240
           | || : |:|:  :||  |  : :::::||| ||:::: |||| : | :  ||:
Db     543 TTTTTTTAWTAWAWWTWAAAWAATWAWWWWWAAAAATWWWWTWTTTTTWATTWATAAWTT 484

Qy     241 AGACTCAATATCTTTACTCTTTCATGAGGACATTGTG 277
           | :|:|| :::  : : : :|||| :  :|:::
Db     483 TWTTTWWAWATTWWKKKKKKAWADTGAGARATKTDKK 447


RESULT  10
CNS00Z0C
LOCUS       CNS00Z0C     1101 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN01C19 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL097062
VERSION     AL097062.1  GI:5608673
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Plasmid Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES              Location/Qualifiers
     source           1. .1101
                      /organism="Drosophila melanogaster"
                      /plasmid="pBeloBAC11"
                      /db_xref="taxon:7227"
                      /clone_lib="DrosBAC"
                      /clone="BACN01C19"
                      /note="end : T7"
BASE COUNT      354 a    173 c    149 g    278 t    147 others
ORIGIN


  Query Match             6.1%;  Score 52.6;  DB 219;  Length 1101;
  Best Local Similarity   29.7%;  Pred. No. 0.051;
  Matches   66;  Conservative   69;  Mismatches   87;  Indels    0;  Gaps    0;
```

```
Qy     419 GACTTCCCAGACAACGTAAGTGTGATAAAAATCTAAAACAAGAGAATTGGCATAAGTTGG 478
           :|:  |: | ||:: : || |||   |||||  :|  ||:|| | |:|   | ||: ||
Db     867 VASSSCSGAMACRMASMAAATGTTKAAAAAAAMTTTAAMAAAAMAMTAAACTTAMTTTAT 926

Qy     479 TGAATGTTTATTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATC 538
             | |  ||    |:::::::||   |::     | |||  ||  : | |: : ::|:|
Db     927 MAACTANTTTAAAAMMHMHMMAAAAAAWWAAAAAAAAAAATTTTTYTTTTWAWAHWTWTA 986

Qy     539 AACGAATCTGCCAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAA 598
           :   :| | |  |  ||      :: :: : :|: ::: ::|:  |:  ||| ::||:
Db     987 WCTAYMTTTYTCTMTMTCTAACAAMMYTMMAHMHAMMHWWTMWAMACAWATAAAWMKCAM 1046

Qy     599 GAACTCATAAAAACCCACACAATGTGAAGCTCTGTTATAAAT 640
           ::  :: ::::: ::::::::|  : :| : |: ||::::::|
Db    1047 MMMAMMMMMMMMMWMMMMMMMACAHAMAMMMCWTTTMYMMMT 1088


RESULT  11
CNS014J2/c
LOCUS       CNS014J2     1025 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN11L11 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL104216
VERSION     AL104216.1  GI:5615827
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Plasmid Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1025)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1025
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN11L11"
                     /note="end : T7"
BASE COUNT      148 a     88 c      6 g     522 t     261 others
ORIGIN


  Query Match             6.1%;  Score 52.4;  DB 219;  Length 1025;
  Best Local Similarity   33.5%;  Pred. No. 0.057;
  Matches  177;  Conservative   83;  Mismatches  267;  Indels    1;  Gaps    1;

Qy     187 TGTCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGGAGACTC 246
           ||: ::|   |||  |  |:| ||::|| |  |     |     |:   ||:| :| |
Db    1017 TGDARKAAAAAAAAAAAAADAAAAWDAAAAAAAAGWGTRAAAATKTAAAAWATRATAWGT 958

Qy     247 AATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGGG 306
             ||| |||  | ||:    | |         | ||: ||||     | |||       |
Db     957 KKTATGTTTKTTTTTWATGWATGTGKRATGWAGAAWGGTTTGTTGDAAAATGKRWAWATG 898

Qy     307 AACCACCTCTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCAT 366
             |   |||    : : |  |   ||   | ::    ||| :: :|   : | | ||
Db     897 TTGTAKTRTTGGGDRARADAAAAGGTGTGATWAGWRAAAATTTRWKTRTTAARTGGTGAT 838
```

```
Qy     367 CAGTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACTAAAGGCTGACTTCCC 426
           |:: | ||||       | | :|| |  |||  :    ::||  |||:    |
Db     837 AARWAAAAAAAAAAAAGWKAGARKGGAGAAAAATGKGGTGWRAAAAAAAKAAAAAAAGKTG 778

Qy     427 AGACAACGTAAGTGTGATAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTT 486
             | || :   :|:||   ||||    |   : |: |::    :   :|:::|||  :|
Db     777 GAAAAAAKRTTRTRTGTGTAAAAAAATATGTRTGRTARWATAAWATTRTKKKTGATGRTA 718

Qy     487 TATTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATC 546
           : : :|| ||  || | :  |  :|  ||| |  |:     |   |  :  |   | |
Db     717 ARAWAWAAAATATAAATAWAYGAAKTGRAAAAAAAARAAAAAAATGTRAWAWATAWATTT 658

Qy     547 TGCCAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCAT 606
           ::   ||   |::| |: |: |:| ||| |:|||||:  ::  ||:: | :| :: |  |
Db     657 KKATTGTGAATWKGTTKGTKKAWAAAAAAAAWAAAAAAWAKWRWTAAWKKTARAAWDGTKKT 598

Qy     607 AAAAACCCACACAATGTGAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTA 666
           :       | : |  | |  : : :|||: ::::       |   || |   || |||
Db     597 WKTKTTTRAAWAAKGGRGTGKGKAKGTTR-RDWWAAAAAAAAAAAAAAAAAAAAAAAGGG 539

Qy     667 TCTACATAAGCAGAAGGAAGAGAAATGAAATACTCATTTTATTGAGTT 714
           |  | | ||: : ||: :||: ||| :| :||     | : :|: ||:
Db     538 TGAADARAAKGDAAARADAGRWAAAARAKRTATAKTGTGKTKTKTGTK 491


RESULT  12
AZ527678
LOCUS       AZ527678      912 bp    DNA             GSS       03-NOV-2000
DEFINITION  ENTBN23TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   AZ527678
VERSION     AZ527678.1  GI:11079801
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 912)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 20
            High quality sequence stop: 734.
FEATURES             Location/Qualifiers
     source          1. .912
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
```

```
BASE COUNT      466 a     51 c     81 g    314 t
ORIGIN


  Query Match             6.0%;  Score 52;  DB 245;  Length 912;
  Best Local Similarity   48.3%;  Pred. No. 0.07;
  Matches  145;  Conservative    0;  Mismatches  155;  Indels    0;  Gaps    0;

Qy     481 AATGTTTATTTAAACATCCAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCAA 540
           |||| |||||||  |  || |  |    |  | ||||||    |  | ||   |||||
Db     179 AATGAATATTTAACATTGAAAATGTTGATTTATTTAATATTTTGTTCAACATAAAATCAA 238

Qy     541 CGAATCTGCCAGTTGCTTTGCTGATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGA 600
             || |    | ||   |     ||  ||  ||   |||| ||  |   ||  |||||||
Db     239 TAAAACATATATTTTGATAATATATAAATGAAATATTAAAGAAATATACAATACTCAAGA 298

Qy     601 ACTCATAAAAACCCACACAATGTGAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAA 660
           ||| ||||||||  |   ||    | |      ||||||   |    ||   ||| | ||
Db     299 ACTAATAAAAACAAAATAAAAAGAATGAATGAATATAAAATTGATAAATAATAGAAGAAA 358

Qy     661 GAAGTATCTACATAAGCAGAAGGAAGAGAAATGAAATACTCATTTTATTGAGTTGGCCCC 720
           ||| ||   | | ||  | ||  ||   | || |||      |       |       |
Db     359 GAAATAGAAAAACAACGAAAAACAAATAATATAAAAAGAATAAAAATATATAAAAGAATA 418

Qy     721 CACTGTATGTGGCTGGTATTTATGAAGGTGATGACCCAGGAAGAAATTGTAAACTATAAA 780
            |    | || |  |     ||| ||  | || |  ||  ||||  |  ||||  |  ||
Db     419 AAAATAAAGTAGAAGAAGAATATAAAAATAATAAATCACAAAGATGTAATAAAAGAAGAA 478


RESULT  13
CNS00EO7
LOCUS       CNS00EO7     1101 bp    DNA             GSS       04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069440
VERSION     AL069440.1  GI:4949583
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR29P01"
                     /note="end : TET3"
BASE COUNT      366 a     66 c    104 g    351 t    214 others
ORIGIN
```

```
  Query Match             6.0%;  Score 52;  DB 219;  Length 1101;
  Best Local Similarity   37.3%;  Pred. No. 0.071;
  Matches  107;  Conservative   44;  Mismatches  136;  Indels    0;  Gaps    0;

Qy       1 TGTTAATGAAGATAAATATTTTTGTTTTTTCGCTGTTCTAAACCTAGGGTTACAAGAAGT 60
           | ||| : :| ||||| |:  ||   ||||   | || || |  :|  |::| || || :
Db     561 TTTTATWWWATATAAAAAWAATTTAATTTTTWATTTTATATATTWAAWGWWAAAAAAAWW 620

Qy      61 AATTTATCTGGAGCTAACAAATACTTTATTTTACCTTTTTATTTGCAAGTAGTTTATGTT 120
           ||  :: :  |  | | |: |: : | ||  |  : |||||||  :   : |:| |  :
Db     621 TATAAWHTWTAAATTTATAWTTWAWATTTTAAATAWATTTATTTAAWTTAWWTWTTTAAW 680

Qy     121 CAATTCTAATTTAATGTATATTAAAAATTCCTCTGCAAATATGTGAGGAGGGACCTCATA 180
            : ||  :  |||| ::: |||:||||:     |  :||||   |  |          |
Db     681 AWTTTAAWTATTAAARWWAATTWAAAAWAAAAWTTTWAATAAWAAATAAAAATAAATTTT 740

Qy     181 AAATATTGTCATATGGAAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAAGG 240
              ||||   ::|:  |||       ||:| :|:  :||:|       | :   ||:|
Db     741 TTTTATTTAAWWAWTTAAAATTAWTWATWAAWAWTTWTAWAAAAAAAAATNWTATAAWATT 800

Qy     241 AGACTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTT 287
           :     |||:| | :: |  :  || | :  | | : |  : | |||
Db     801 WWKAATAATWTATAWWTTTAWATATWATRTAAATTWTAATWYTTTTT 847


RESULT  14
AZ548198/c
LOCUS       AZ548198      853 bp    DNA             GSS       14-NOV-2000
DEFINITION  ENTFW05TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   AZ548198
VERSION     AZ548198.1  GI:11171567
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 853)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence start: 18
            High quality sequence stop: 838.
FEATURES             Location/Qualifiers
     source          1. .853
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
```

Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999)."
BASE COUNT 222 a 127 c 68 g 436 t
ORIGIN

Query Match 6.0%; Score 51.8; DB 245; Length 853;
Best Local Similarity 50.2%; Pred. No. 0.077;
Matches 128; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

Qy 444 TAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTATTTAAACATCCAATT 503
Db 425 TTAATATATTAAACATAAACAACATTTCAGAATAAAGAATGAATATTTAACAATGAAAGT 366

Qy 504 CATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATCTGCCAGTTGCTTTGCTG 563
Db 365 GTTGATTTATTTAATATTTTGTTCAACATAAATTCAATAAAACATATATTTTGATAATAT 306

Qy 564 ATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAAAAACCCACACAATGT 623
Db 305 ATAAATGGAATATTAAAGGAATATACAATACTCAAGAACTAATAAAAAAGAAATAAAAAG 246

Qy 624 GAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATCTACATAAGCAGAAGG 683
Db 245 AATGAATAAATGAAGATAAATTTGATAAATAATACAAGAAAGAAATATAAACGACAAAGA 186

Qy 684 AAGAGAAATGAAATA 698
Db 185 AAAAGAAATAATAGA 171

RESULT 15
AZ686179/c
LOCUS AZ686179 868 bp DNA GSS 14-DEC-2000
DEFINITION ENTJI41TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
ACCESSION AZ686179
VERSION AZ686179.1 GI:11823325
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 868)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica HM1:IMSS sheared DNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 74
High quality sequence stop: 779.
FEATURES Location/Qualifiers
source 1. .868
/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOS1; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999)."
BASE COUNT 214 a 128 c 82 g 444 t
ORIGIN

Query Match 6.0%; Score 51.8; DB 248; Length 868;
Best Local Similarity 50.2%; Pred. No. 0.077;
Matches 128; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

Qy 444 TAAAAATCTAAAACAAGAGAATTGGCATAAGTTGGTGAATGTTTATTTAAACATCCAATT 503
Db 419 TTAATATATTAAACATAAACAACATTTCAGAATAAAGAATGAATATTTAACAATGAAAGT 360

Qy 504 CATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATCTGCCAGTTGCTTTGCTG 563
Db 359 GTTGATTTATTTAATATTTTGTTCAACATAAATTCAATAAAACATATATTTTGATAATAT 300

Qy 564 ATGCATAGAAAGATAAAAAAGAAAGAAAAGCTCAAGAACTCATAAAAACCCACACAATGT 623
Db 299 ATAAATGGAATATTAAAGGAATATACAATACTCAAGAACTAATAAAAAAGAAATAAAAAG 240

Qy 624 GAAGCTCTGTTATAAATGGGTGCCATGTAAGATGGAAGAAGTATCTACATAAGCAGAAGG 683
Db 239 AATGAATAAATGAAGATAAATTTGATAAATAATACAAGAAAGAAATATAAACGACAAAGA 180

Qy 684 AAGAGAAATGAAATA 698
Db 179 AAAAGAAATAATAGA 165

Search completed: November 20, 2001, 15:31:57
Job time: 3966 sec